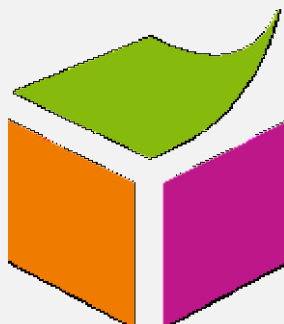


Linked Data Scholarship: Modeling and Interpretation in the Digital Humanities

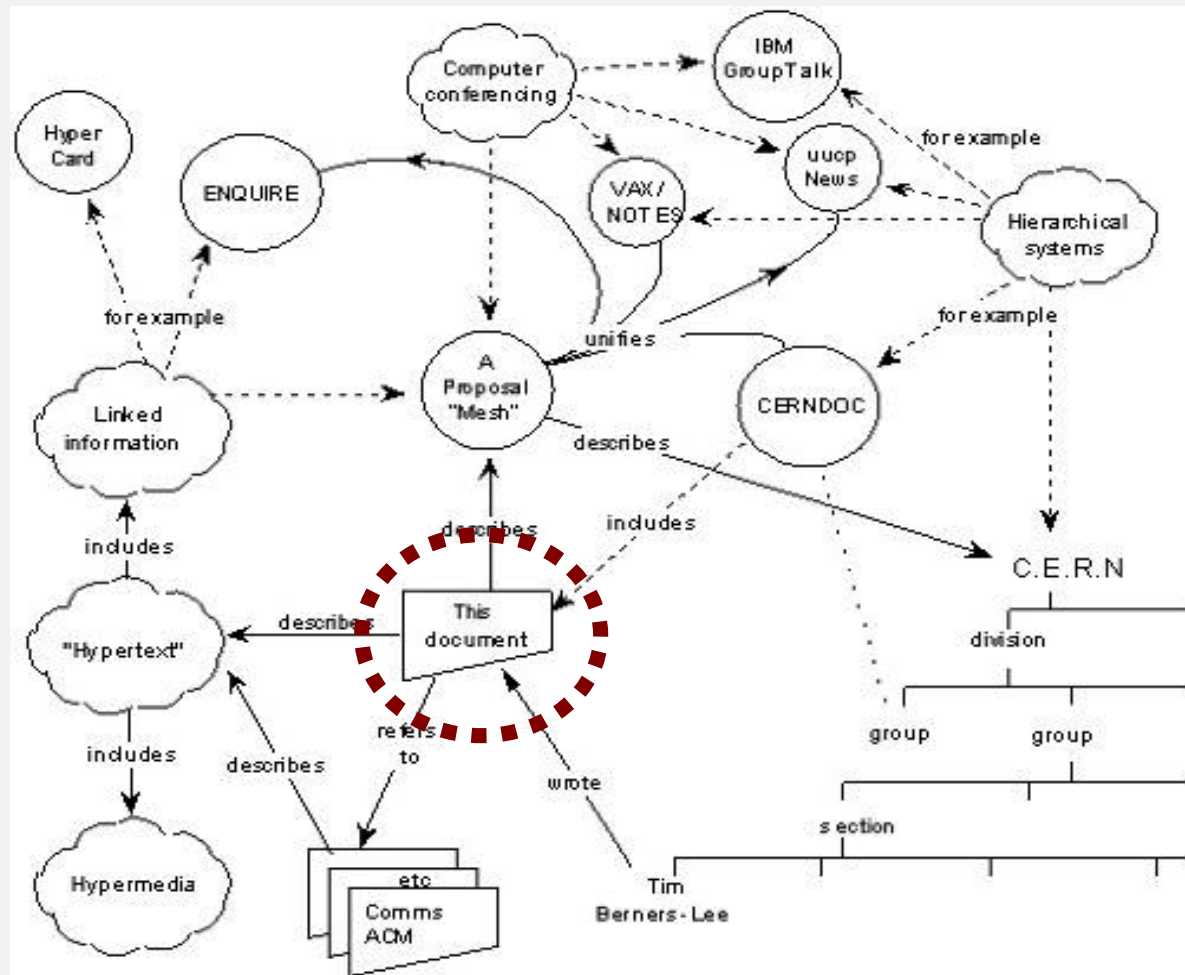
Prof. Dr. Stefan Gradmann (KU Leuven)

Universidad Carlos III de Madrid, 11/07/2013



- An Introduction to RDF (?)
- DM2E
- The Scholarly Domain Model
- What IS a model, anyway?
- The Wittgenstein Incubator

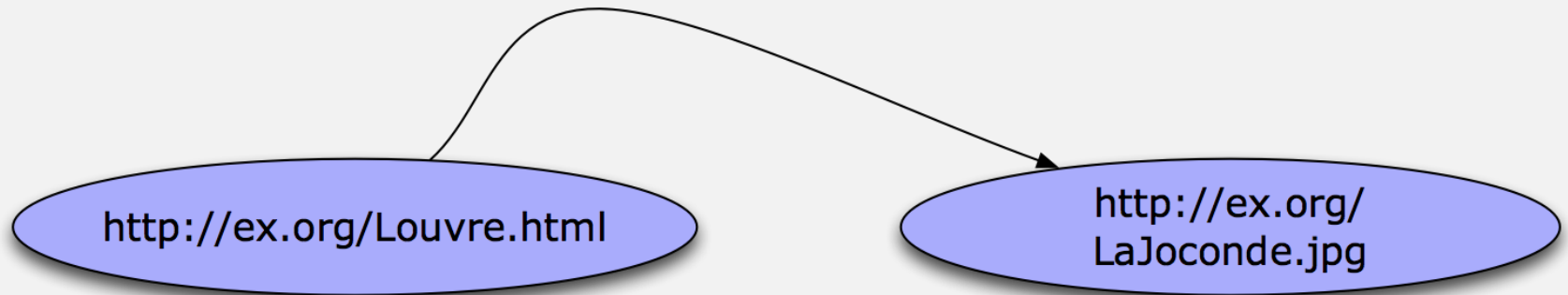
The Web of Documents



Information Management: A Proposal (TBL, 1989)

... twice extended:
• in syntax
• in scope

Resources and Links in the Document Web

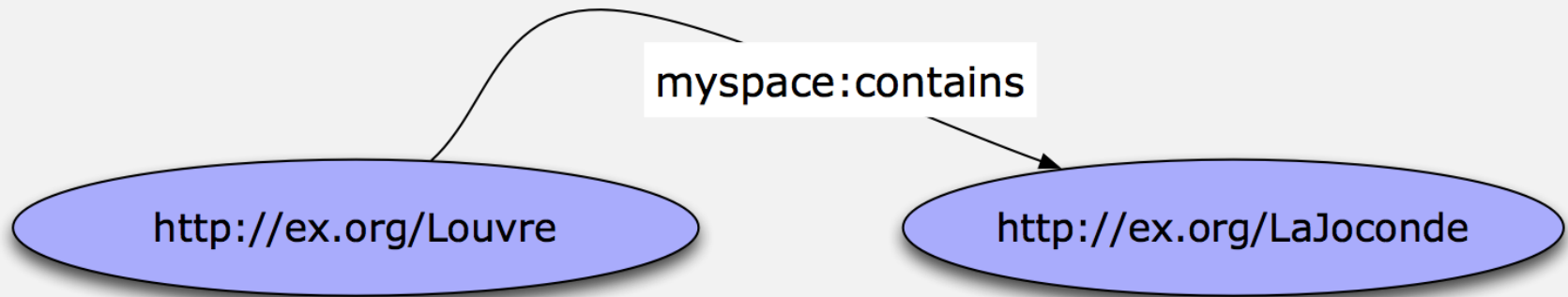


```
<a href="http://ex.org/LaJoconde.jpg">b</a>
```

- We have HTTP URIs to identify resources and links between them – but we are missing a few things!
- What kinds of resources are 'Louvre.html' and 'LaJoconde.jpg'?
 - A machine cannot tell.
 - Humans can: we recognize implied context!
- How exactly do they relate to each other?
 - A machine cannot tell.
 - Humans can: again we recognize implied context!

Syntactically Extending the Document Web (1)

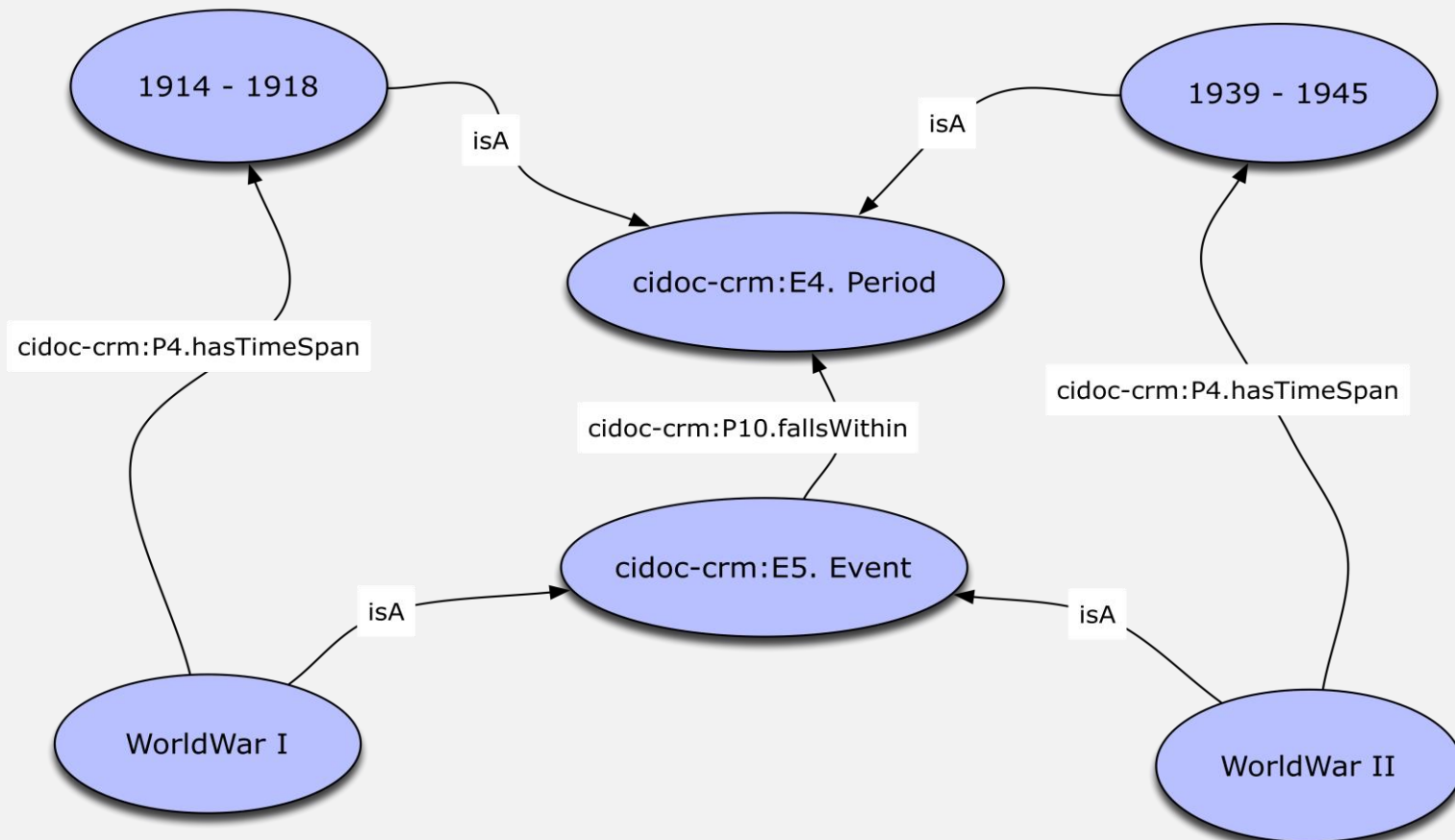
- We add a syntax for making statements on resources: RDF triples



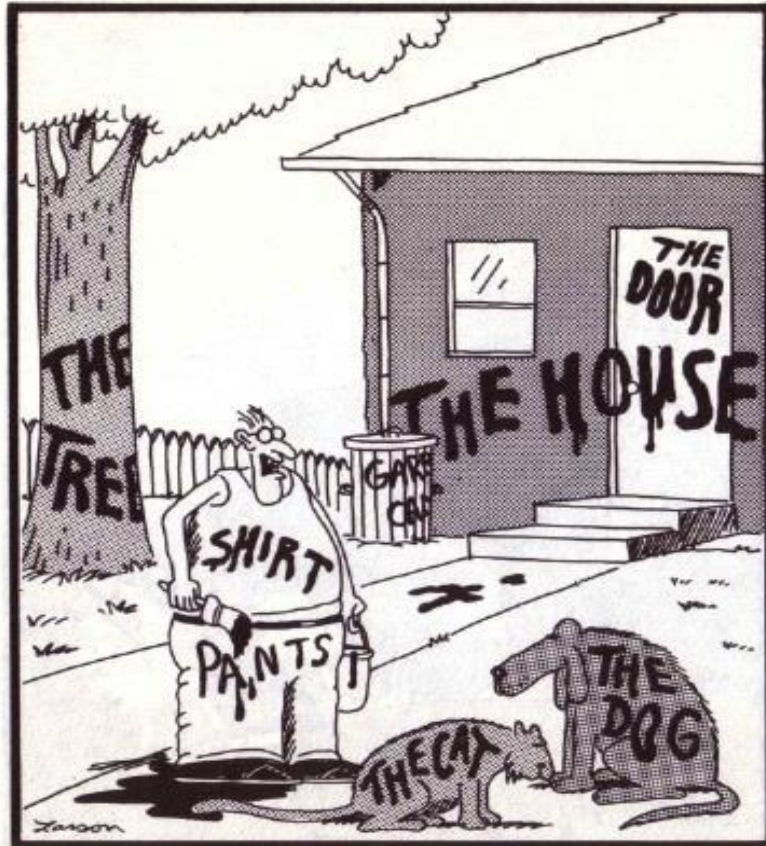
- We add a schema language (RDFS) with elements such as
 - classes ('chair' as instance of chairs),
 - hierarchies of classes and properties (chairs are a subclass of furniture, 'teaches' is a sub-property of 'communicates')
 - inheritance (communication based on language → teaching also is)
 - support for basic inferencing, deterministic logical operations

Syntactically Extending the Document Web: RDF (2)

- And thus are able to establish structures in triple aggregations resulting in lightweight domain ontologies:



Extending the Web in Scope: The Web of Things ... (slightly Mistaken)

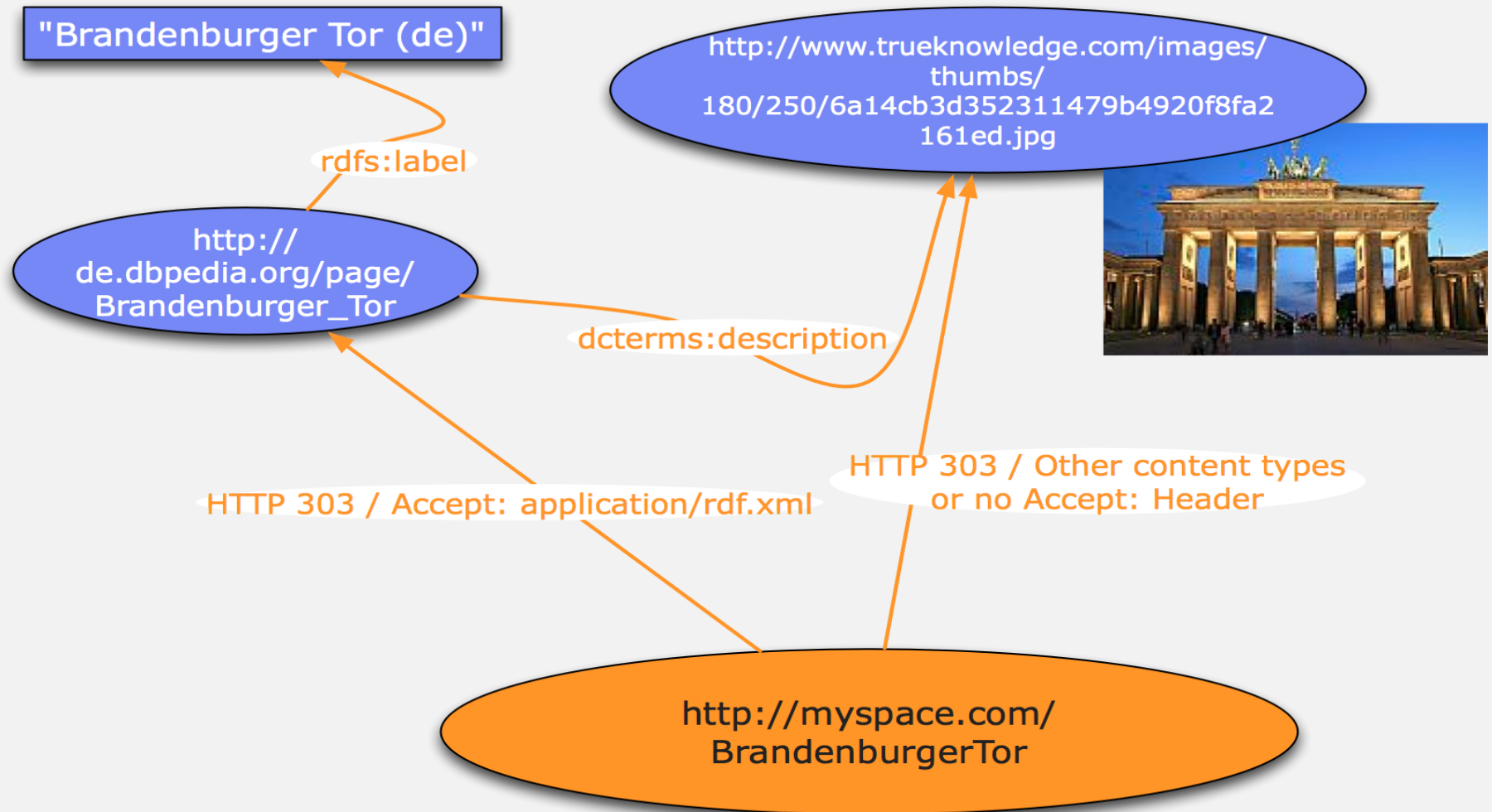


"Now! ... That should clear up
a few things around here!"

What's **wrong**
with this picture?

Taken from Ronald Carpentier's
Blog at

... and the Way we extend the Web in scope to make it a 'Web of Things'



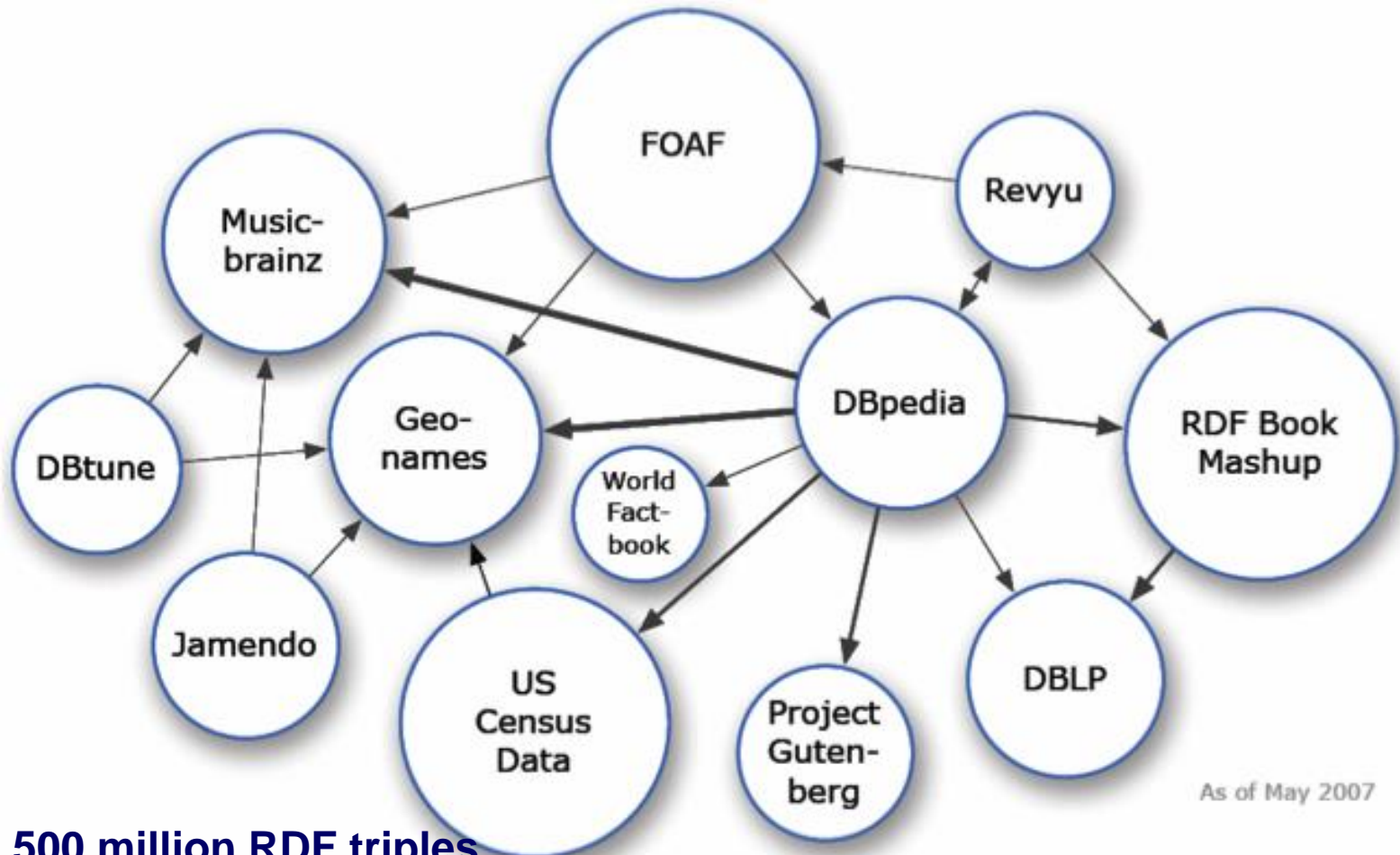
Linked Data essentials

1. Use URIs Standard **Identifiers**
 2. Use HTTP URIs Standard **Pointers**
 3. Serve useful information using SPARQL, RDF standards
 4. Mention URIs of related objects Standards for **Queries**
and **Statements**
- Link** to Context

Copyright © 2008 W3C (MIT, ERCIM, Keio)

[http://www.w3.org/2008/Talks/0617-lod-tbl/#\(4\)](http://www.w3.org/2008/Talks/0617-lod-tbl/#(4))

A few Bubbles: 5/2007

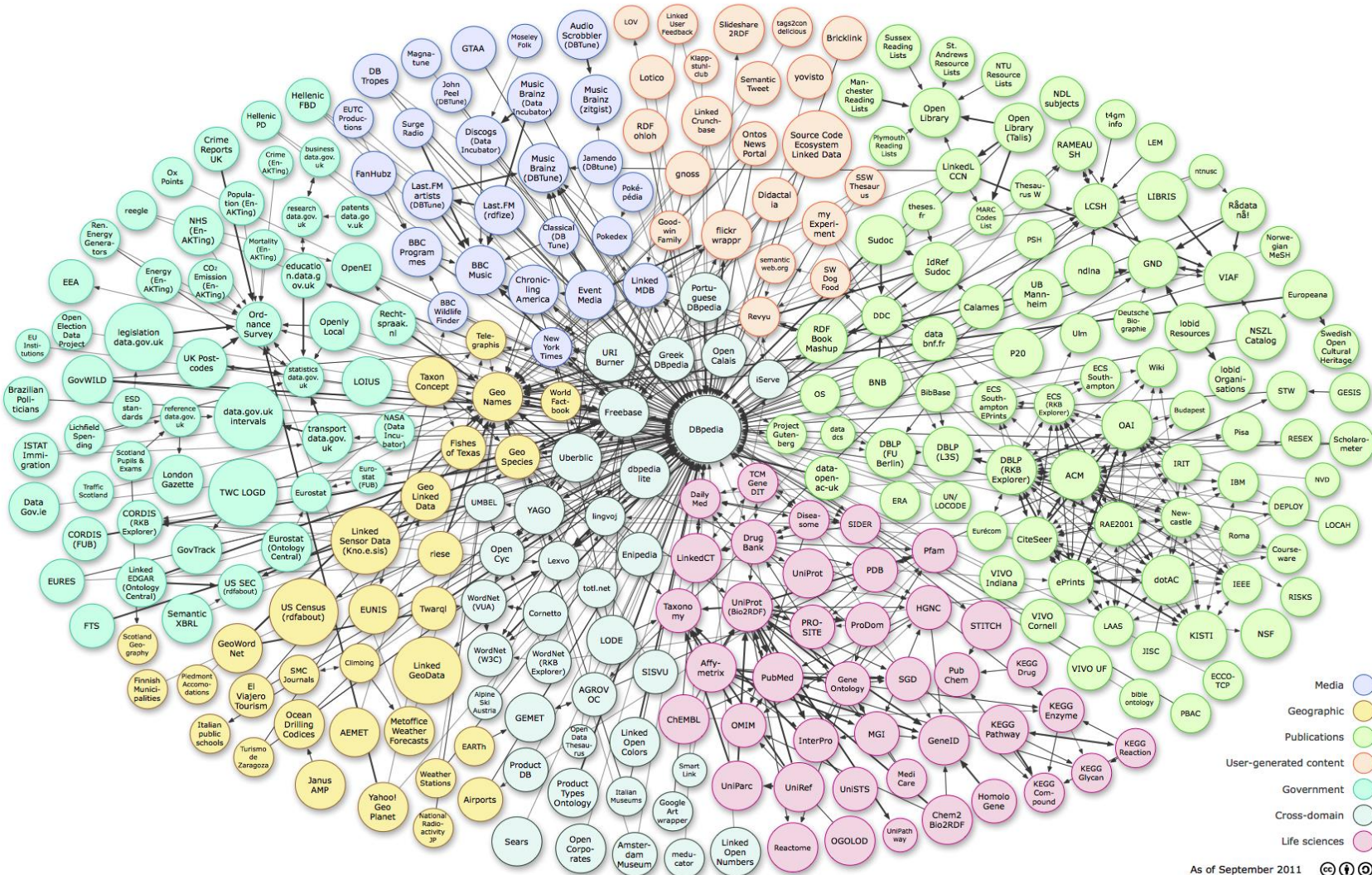


- Over 500 million RDF triples
- Around 120,000 RDF links between data sources

© Richard Cyganiak

And a lot of Bubbles as of last Year

KU LEUVEN



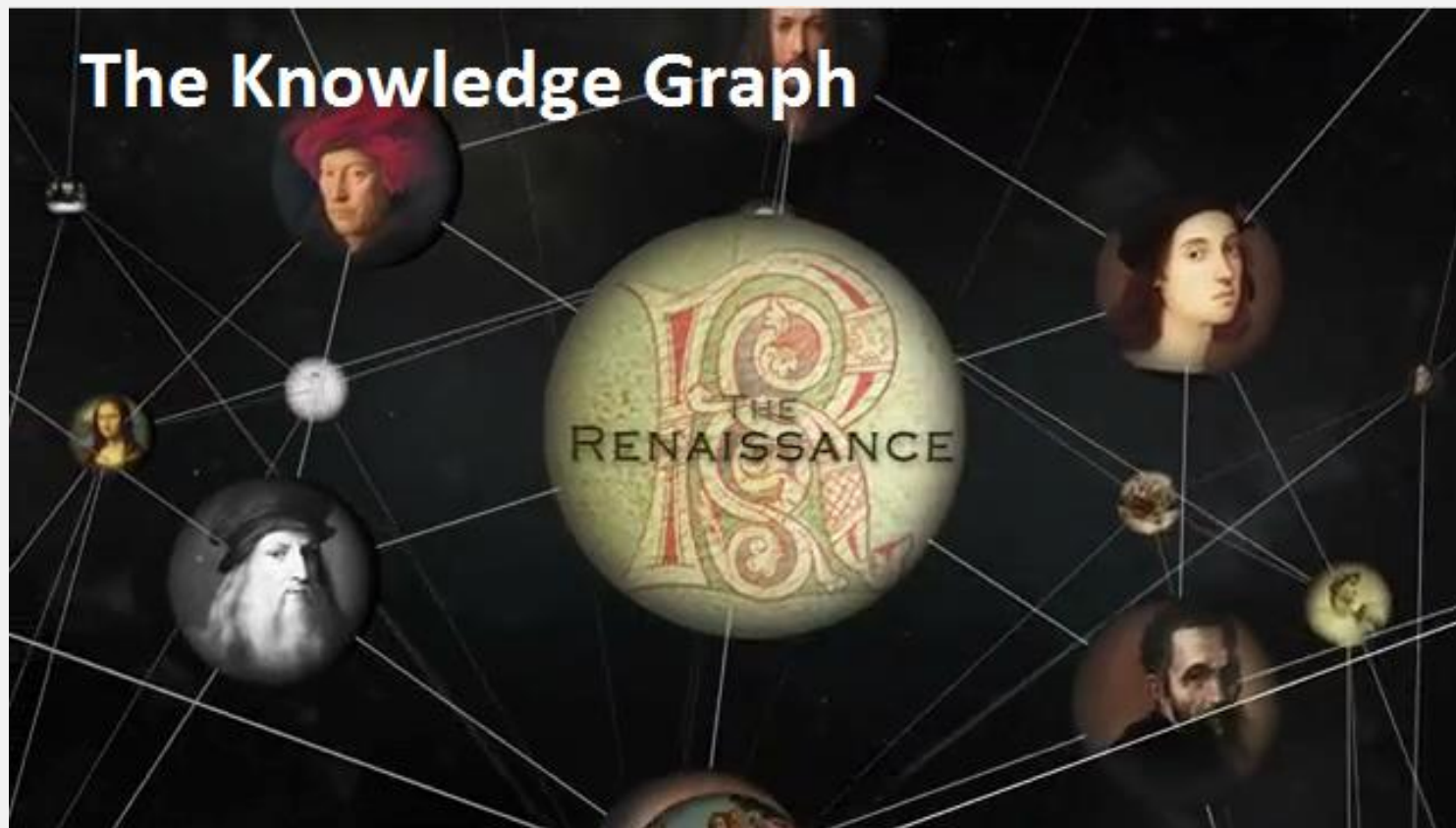
Linked Data Scholarship: Modeling and Interpretation in the Digital Humanities

Stefan Gradmann, Universidad Carlos III de Madrid, 11/07/2013

11

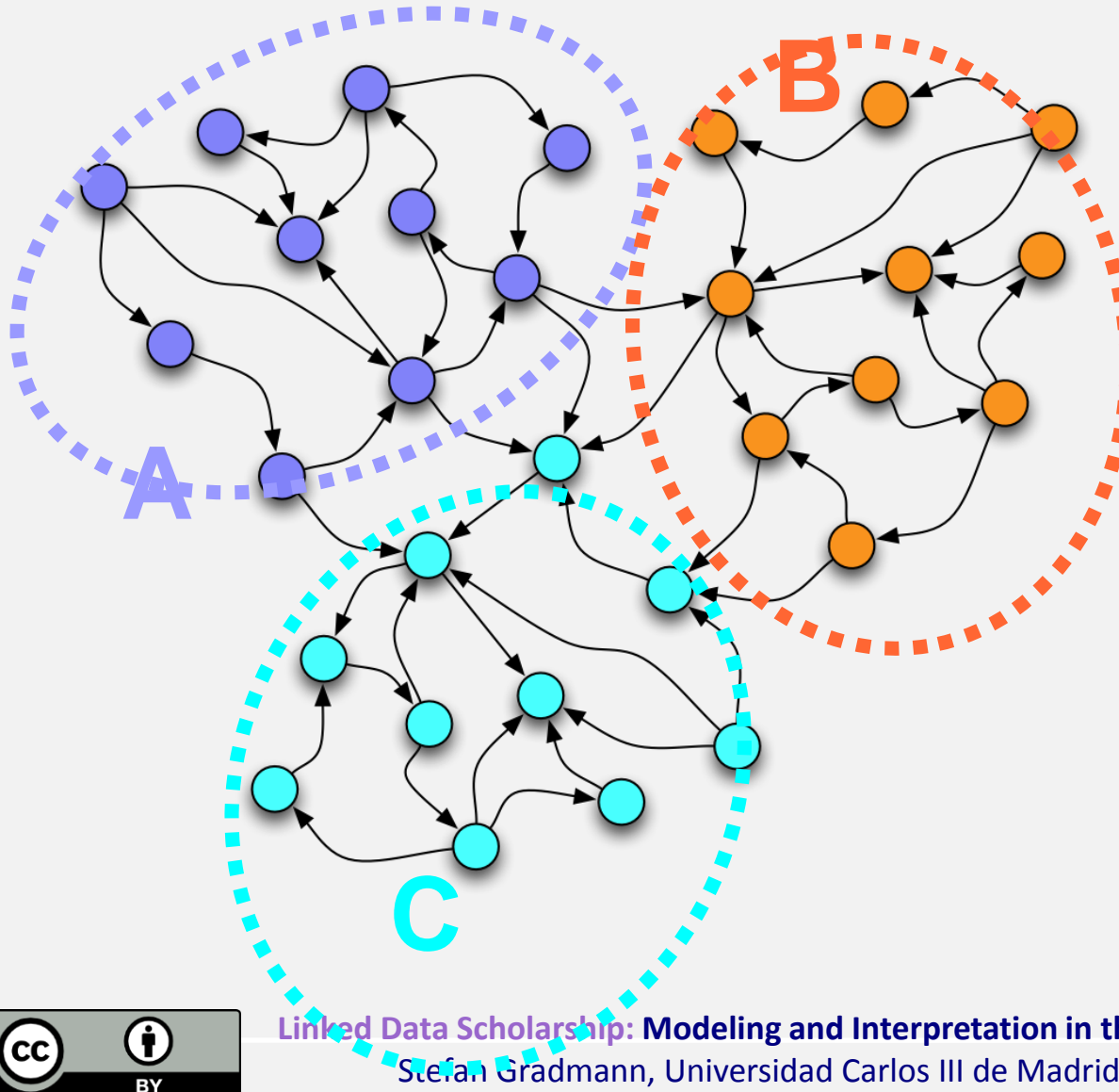
Google entering the Floor

KU LEUVEN



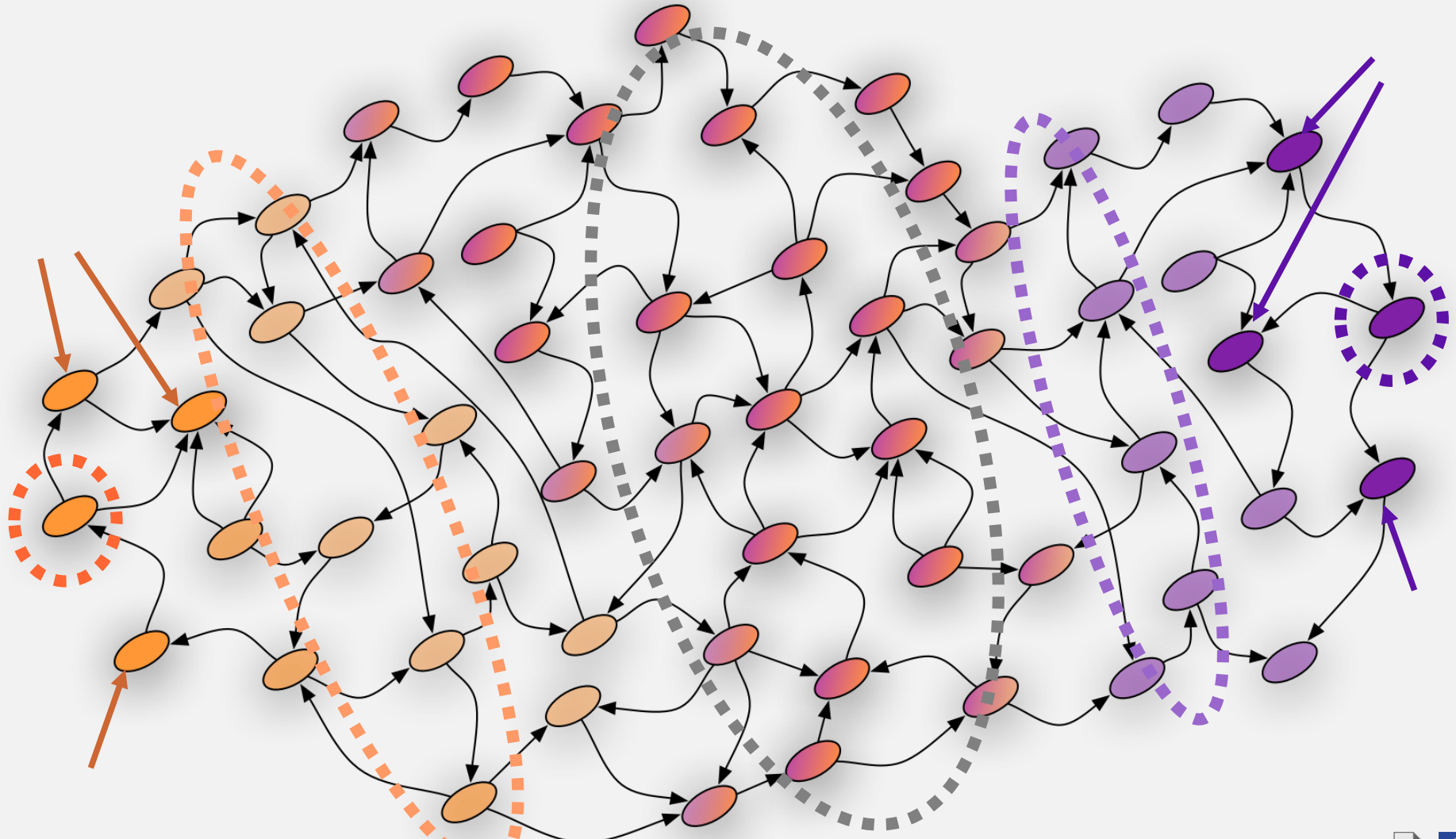
Modelling Object Representations as RDF

Aggregations generates new questions ...



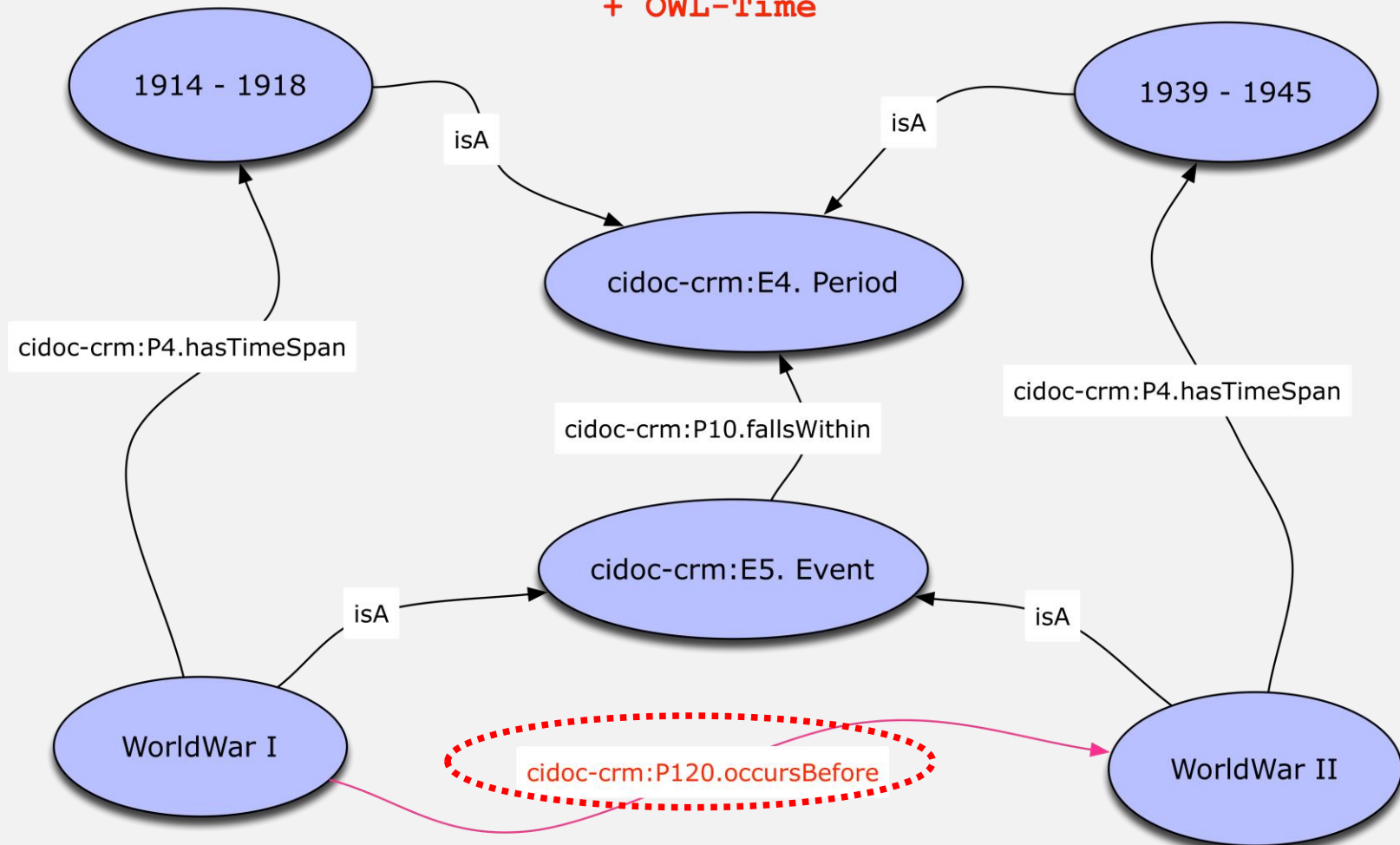
- Where do resource aggregations 'start'? Where do they 'end'?
 - And what constitutes document boundaries??
 - And which node was connected to which one at a given time???
- Provenance,
Versioning,
Authorisation: Named
Graphs

Aggregations and Context: Calculating Closeness



... and new opportunities: Triple Sets and 'Reasoning'

+ OWL-Time





02/2012-01/2015

Co-funded by the EC

Digitised Manuscripts to Europeana (DM2E): Who (1)?

- **Content Providers**

- European Association for Jewish Culture (Judaica)
- Max-Planck-Institut für Wissenschaftsgeschichte (ECHO)
- Österreichische Nationalbibliothek (Google)
- Staatsbibliothek zu Berlin (Kalliope)
- University of Bergen (Wittgenstein)
- CNRS ITEM (Nietzsche)
- National Library of Israel (Judaica)
- Berlin Brandenburgische Akademie (German Text Archive)
- Humboldt-Universität zu Berlin (Polytechnisches Journal)

- **Technology Providers**

- ExLibris (Aleph, MARC sources management)
- Universität Mannheim / Freie Universität Berlin (LoD2, D2R, SILK)
- Max-Planck-Institut für Wissenschaftsgeschichte (ECHO)
- Net7 S.r.l. (Muruca/Pundit)
- National Technical University of Athens (MINT)

Digitised Manuscripts to Europeana (DM2E): Who (2)?

- **Digital Humanities Community**

- Dr. Tobias Blanke (King's College, London)
- Sally Chambers (The European Library / DARIAH-D)
- Prof. Dr. Stefan Gradmann (KU Leuven, Chair)
- Prof. Dr. Gerhard Lauer (Göttingen University)
- Dr. Alois Pichler (UIB)
- Dr. Jürgen Renn (MPIWG)
- Dr. Laurent Romary (HUB)
- Prof. Dr. Susan Schreibman (Trinity College Dublin)
- Dr. Claire Warwick (University College, London)

- **Community Building**

- Open Knowledge Foundation (OKFN)

- **Coordination, Management & Information Science**

- Humboldt-Universität zu Berlin (HUB)

Digitised Manuscripts to Europeana (DM2E): Who (3)?

- **New Associated Partners**
 - Universitätsbibliothek Frankfurt a.M. (UBFFM)
 - Bulgarian Academie of Sciences (BAS)
 - Ontotext
 - Brandeis University
 - Georg Eckert Institute for International Textbook Research (GEI)
 - Joint Distribution Committee (JDC)

Digitised Manuscripts to Europeana (DM2E): What?

- Provide substantial amounts of **digital content** to Europeana with a focus on **digitised manuscripts** (WP1)
- Integrate existing **technical building blocks**
 - from Europeana development
 - as well as from generic LoD oriented development
 - into a **generic production chain** for migrating data from various sources to the EDM as well
 - as for the **contextualisation** of the object representations (WP2).
- Explore usage scenarios of EDM metadata together with object data in a specialised RDF graph based platform for **humanities research** making available specialised visualisation and reasoning environments (WP3).

WP3: Digital Humanities Requirements and Related Engineering - Context

Goal: lower the barriers for **digital content curation** by providing an integrated, flexible, semantic based environment targeted to **digital humanities scholars**

Structured or semi-structured data sources

Compliant web pages with metadata

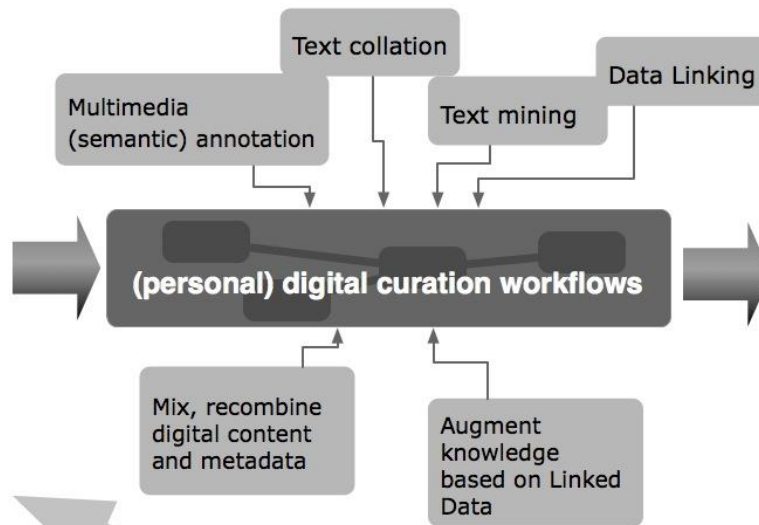
Europeana (EDM)

Partners/third party platforms

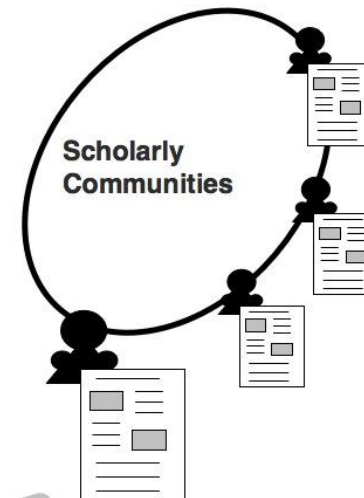
Linked Open Data Cloud

Web of Data

Digital re-mix and curation tool
for humanities scholars and domain experts



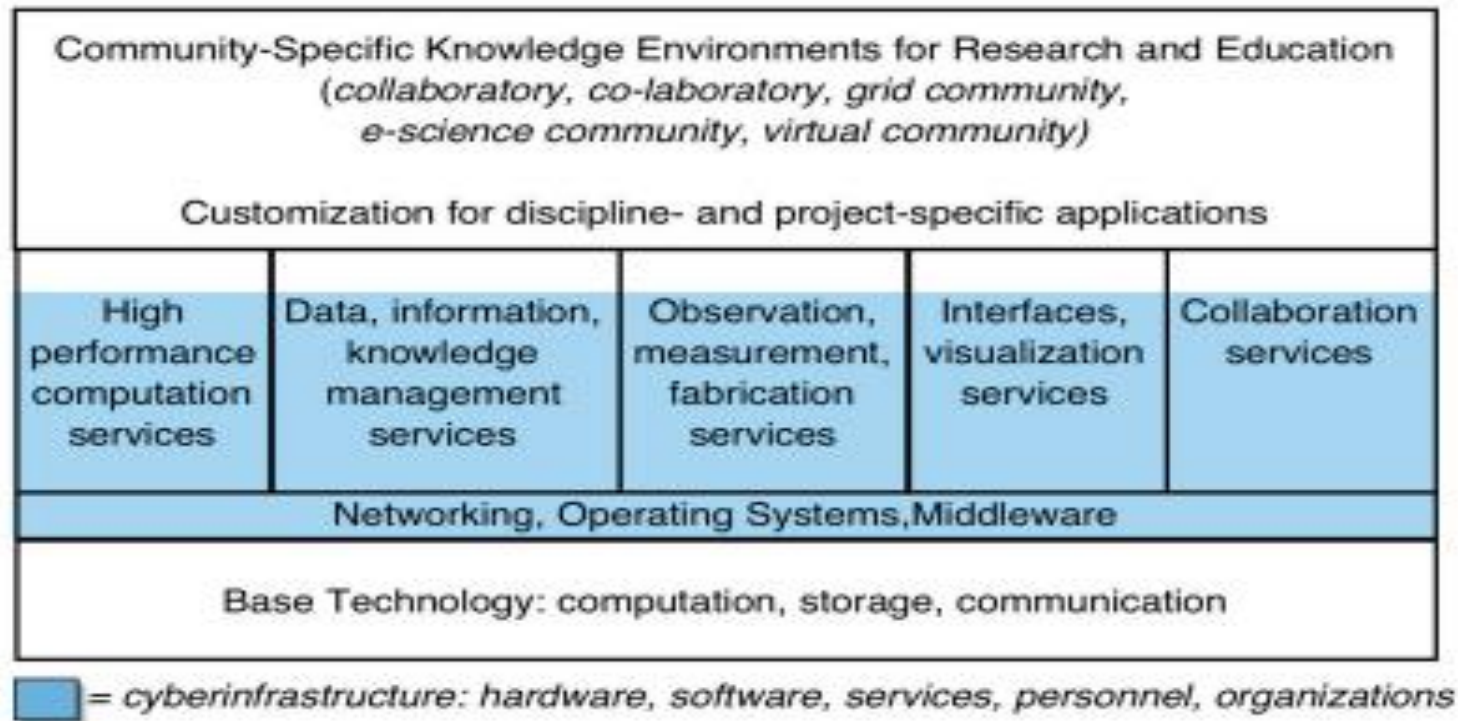
- Custom virtual collections
- Curated, domain specific mashups
- Scientific contributions
- etc..



- Prototype **platform enabling digital scholarship** in combining EDM RDF metadata, digital surrogates and Linked Data ...
- ... building on an **ontological representation of scholarly work** based on a common understanding of its constituents
- ... resulting in a increasingly complex **social semantic scholarly graph** containing RDF statements such as
 - Version_A – isSuccessorOf – Version_B
 - Statement_1 – contradicts - Statement_2
 - Scribe_Y – copiedFrom – Scribe_Z
- ... and which could **feed back richly contextualised EDM** to Europeana!

Beyond Infrastructure: The Scholarly Domain Model

Cyberinfrastructure: Atkins Report (2003)



- “Mother of all infrastructure layer cakes” impacted
 - “Our Cultural Commonwealth”, e-Science (UK), TextGrid, DARIAH
 - With Isidore, Europeana and others being more content oriented and LoD based

Why Beyond Infrastructure?

- We want to move **beyond emulation mode** ...
- ... beyond 'pages' and 'links'
- “**Research infrastructure is not research** just as roads are not economic activity. We tend to forget when confronted by large infrastructure projects that they are not an end in themselves. [...] Infrastructure projects can become ends in themselves by developing into an industry that promotes continued investment. To sustain infrastructure there develops a class of people whose jobs are tied to infrastructure investment.”

Rockwell (2010)

Linked Data Scholarship: Modeling and Interpretation in the Digital Humanities

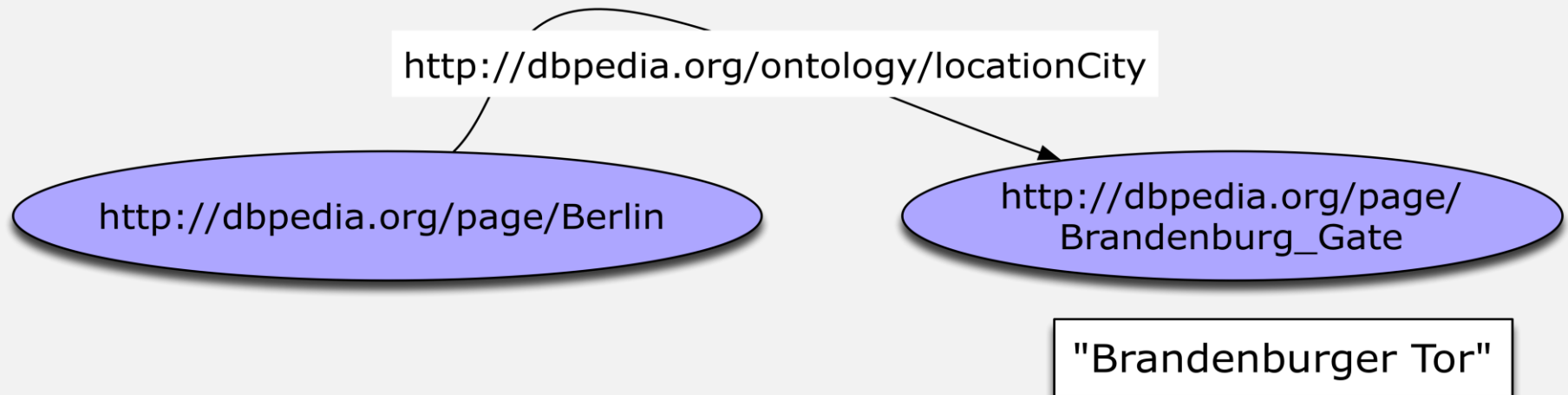
Stefan Gradmann, Universidad Carlos III de Madrid, 11/07/2013

- ⇒ how can we better **understand and model**



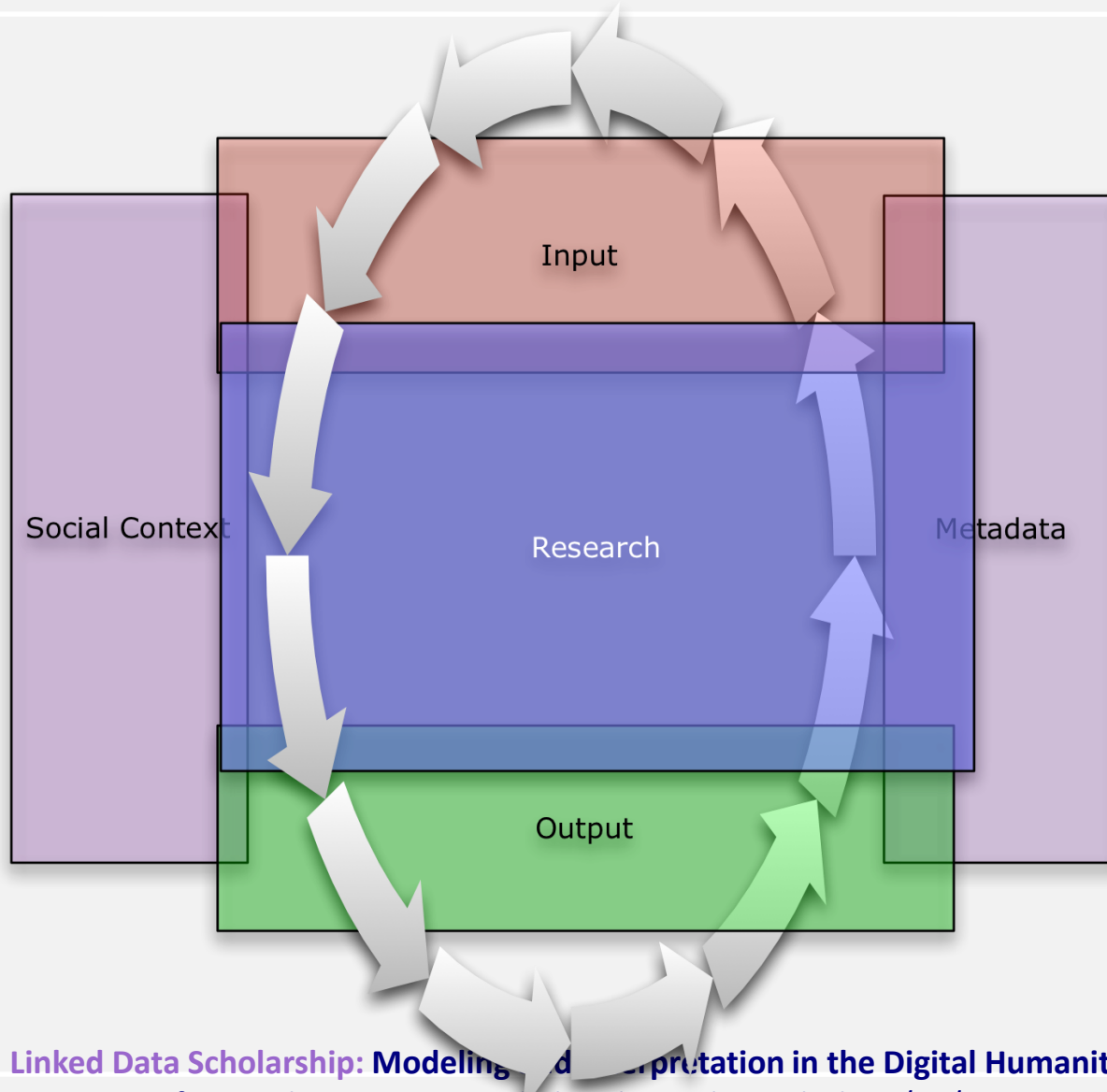
- Unsworth (2000)
 - *discovering, annotating, comparing, referring, sampling, illustrating, representing*
 - as the basis for tool-building enterprises for the Digital Humanities
- Palmer et al. (2009) (“scholarly information activities”)
 - *searching, collecting, reading, writing, collaborating*
- ... Blanke & Hedges (2011), Bamboo (2010), McCarty et al. (2002) Anderson et al. (2010) ...
- Bernardou et al. (2010)
 - CRM activity and event based process model connecting research activities with information objects and propositions, i.e. including argumentation structures

- Typed statements on web resources (triples) and how they relate to each other, e. g.

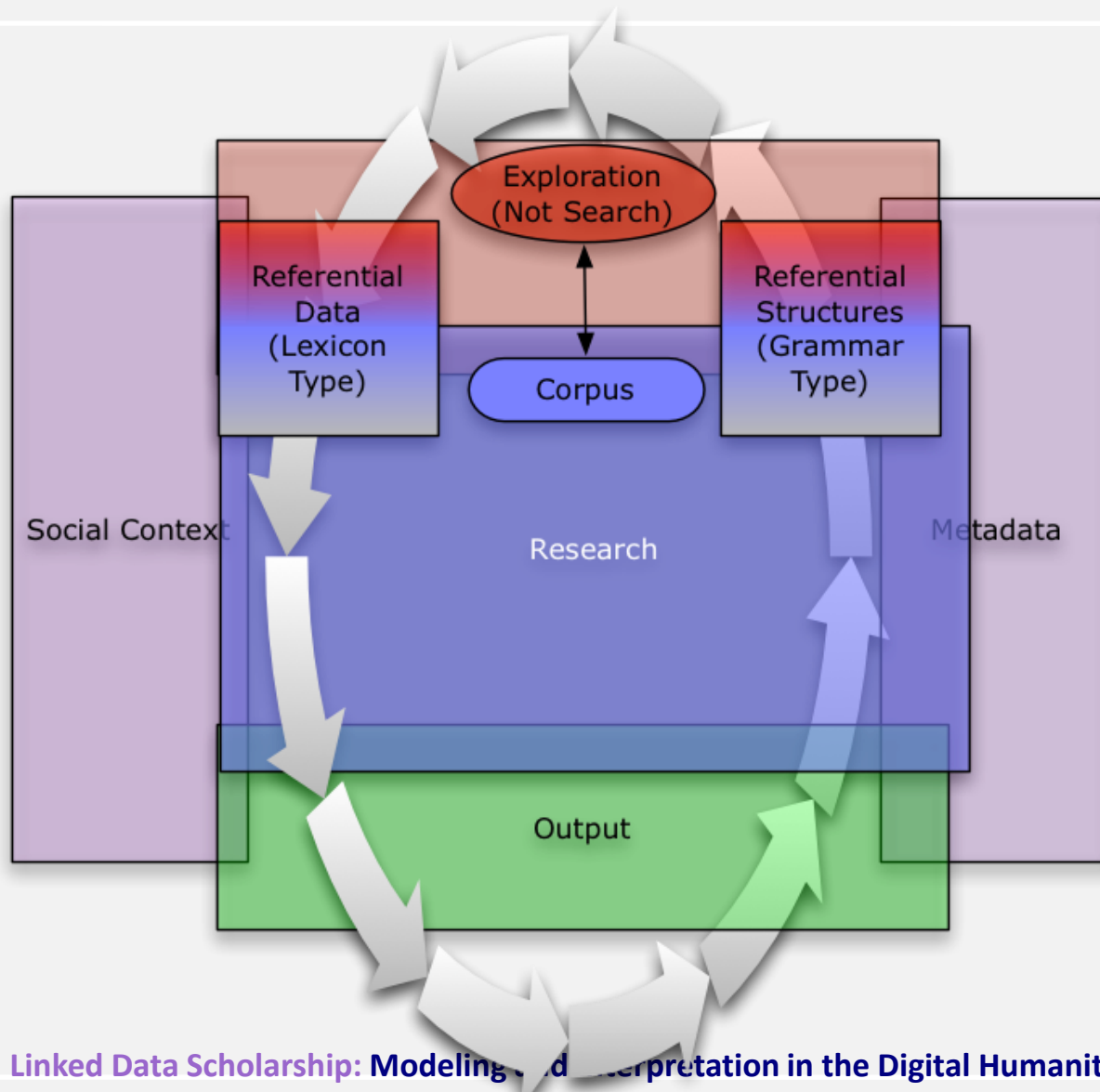


- + RDF Schema (RDFS) language with constructors for sub- and superclasses and -properties including the concept of inheritance
- → simple, deterministic logical operations on triple aggregations (“reasoning”)

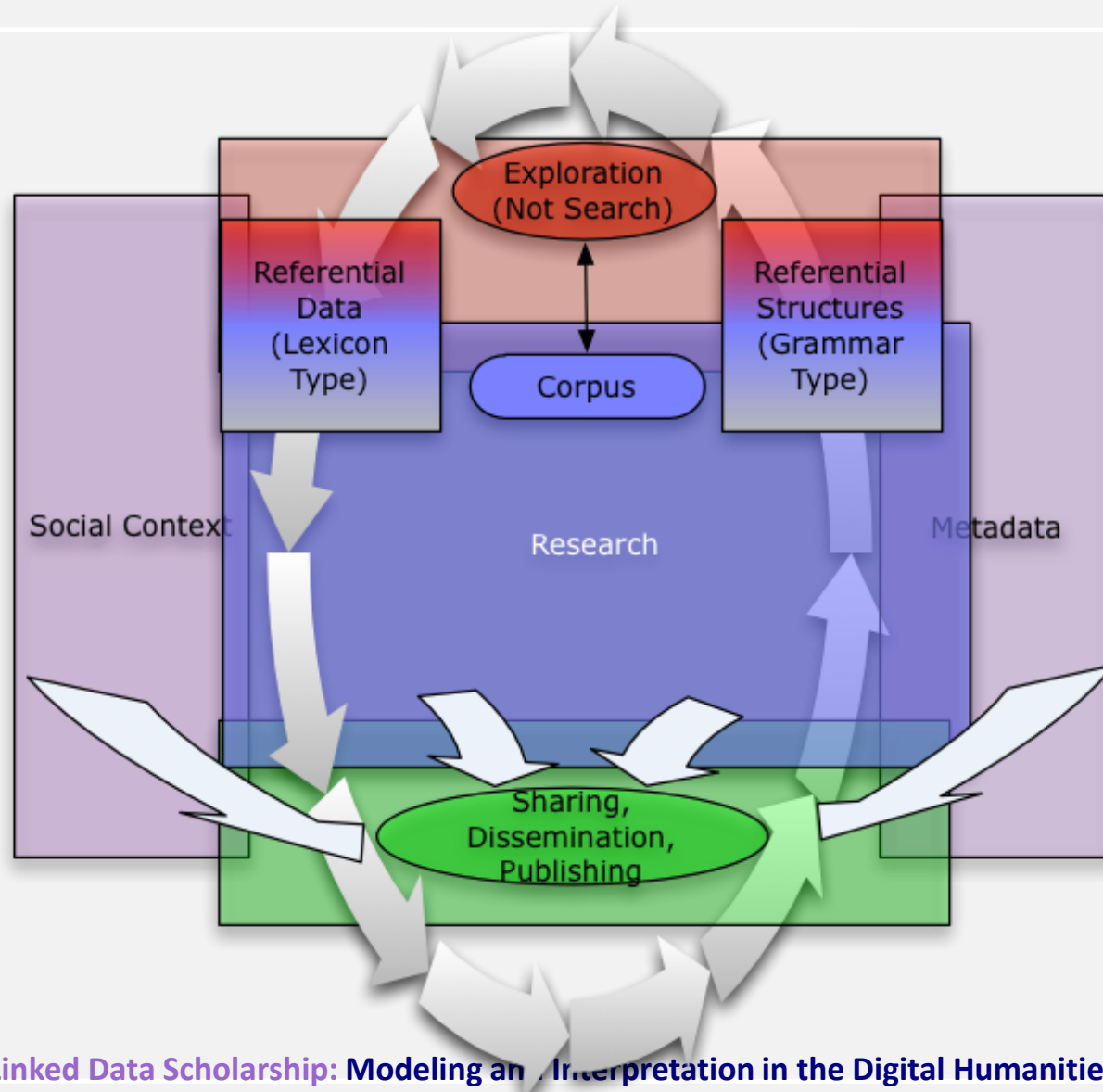
The Scholarly Domain from 10.000 feet above



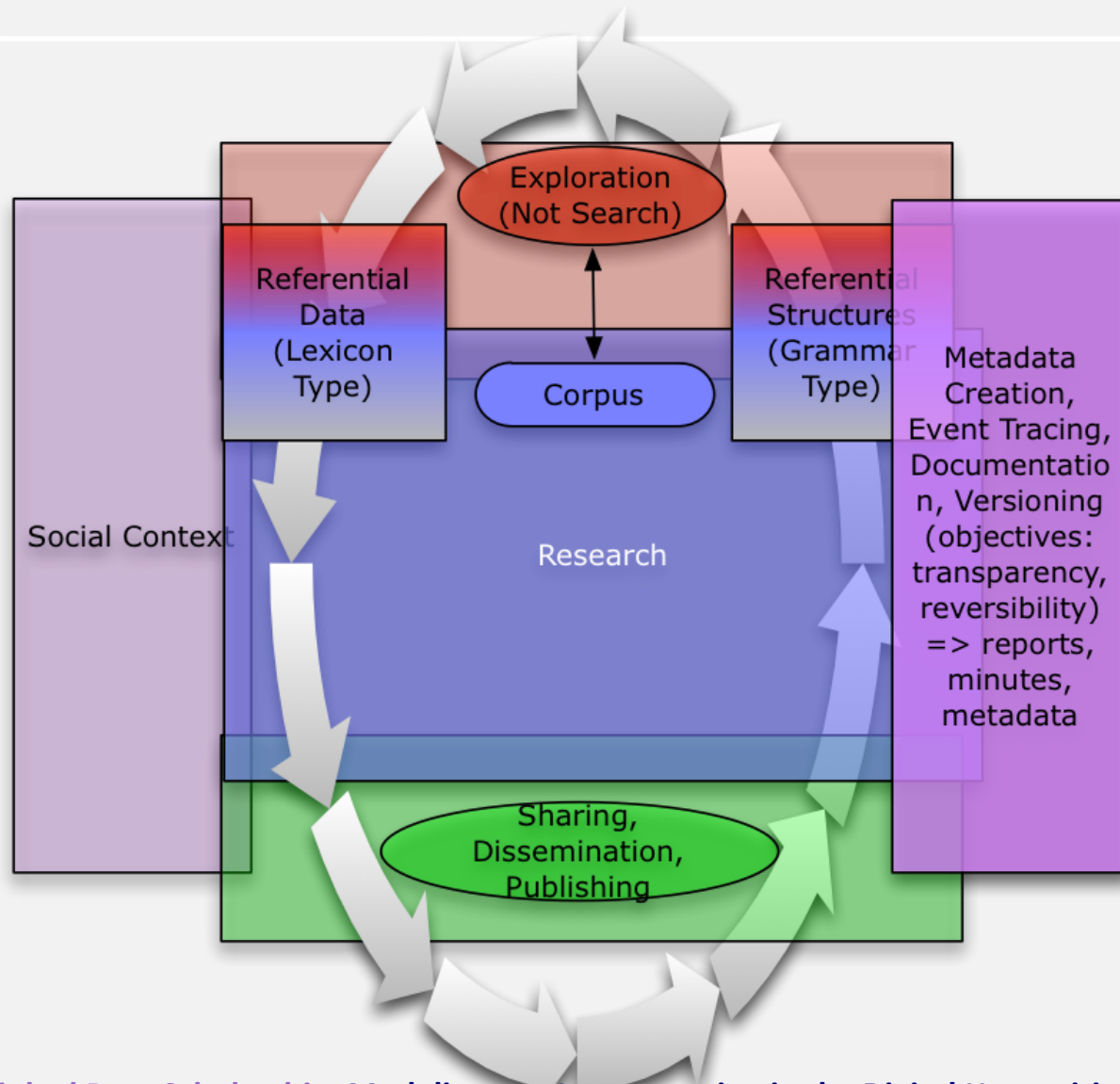
Input Area Details



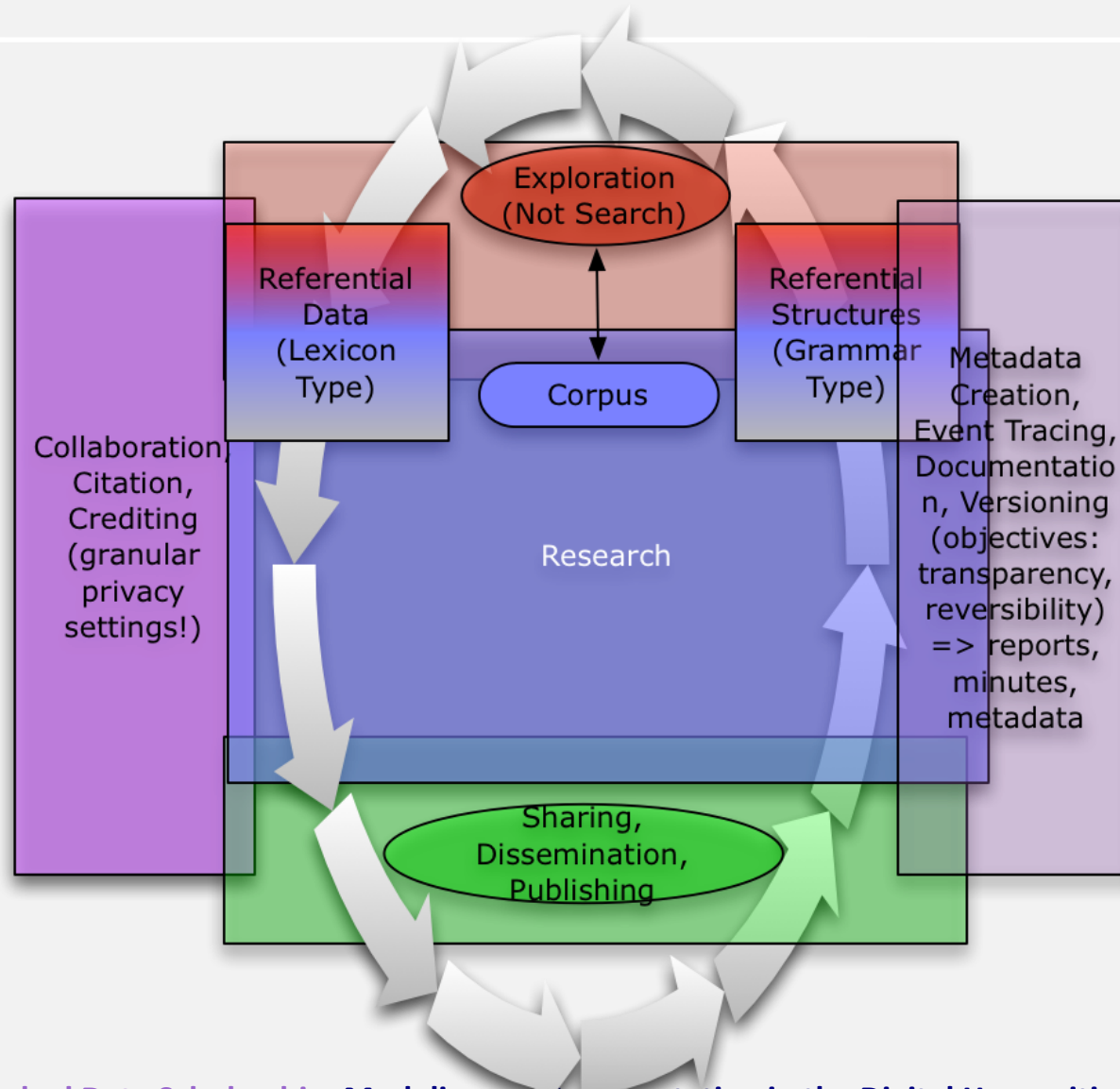
+ Output



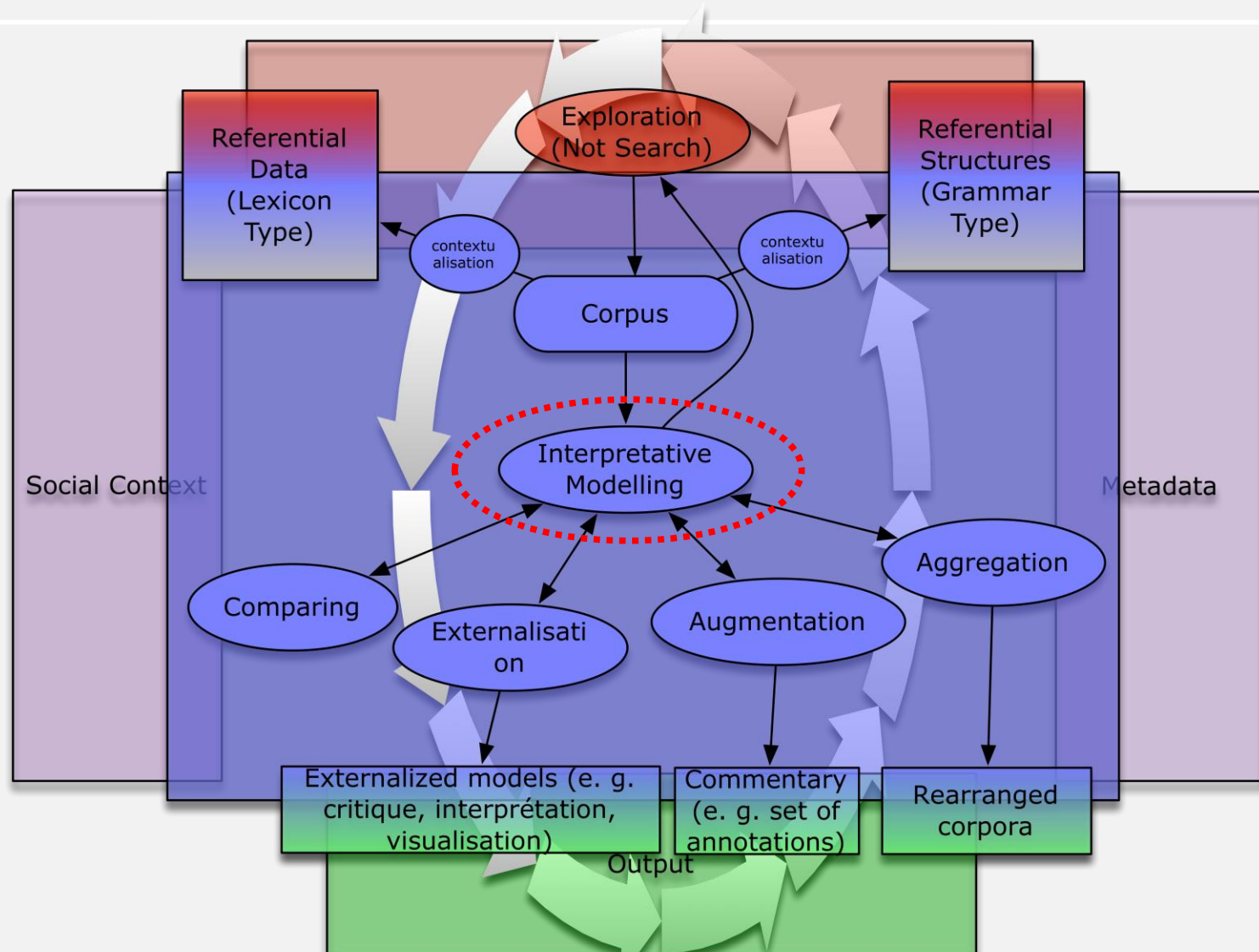
+ Metadata



+ Social Context



Zoom on Research



A normative Model



A Car Model



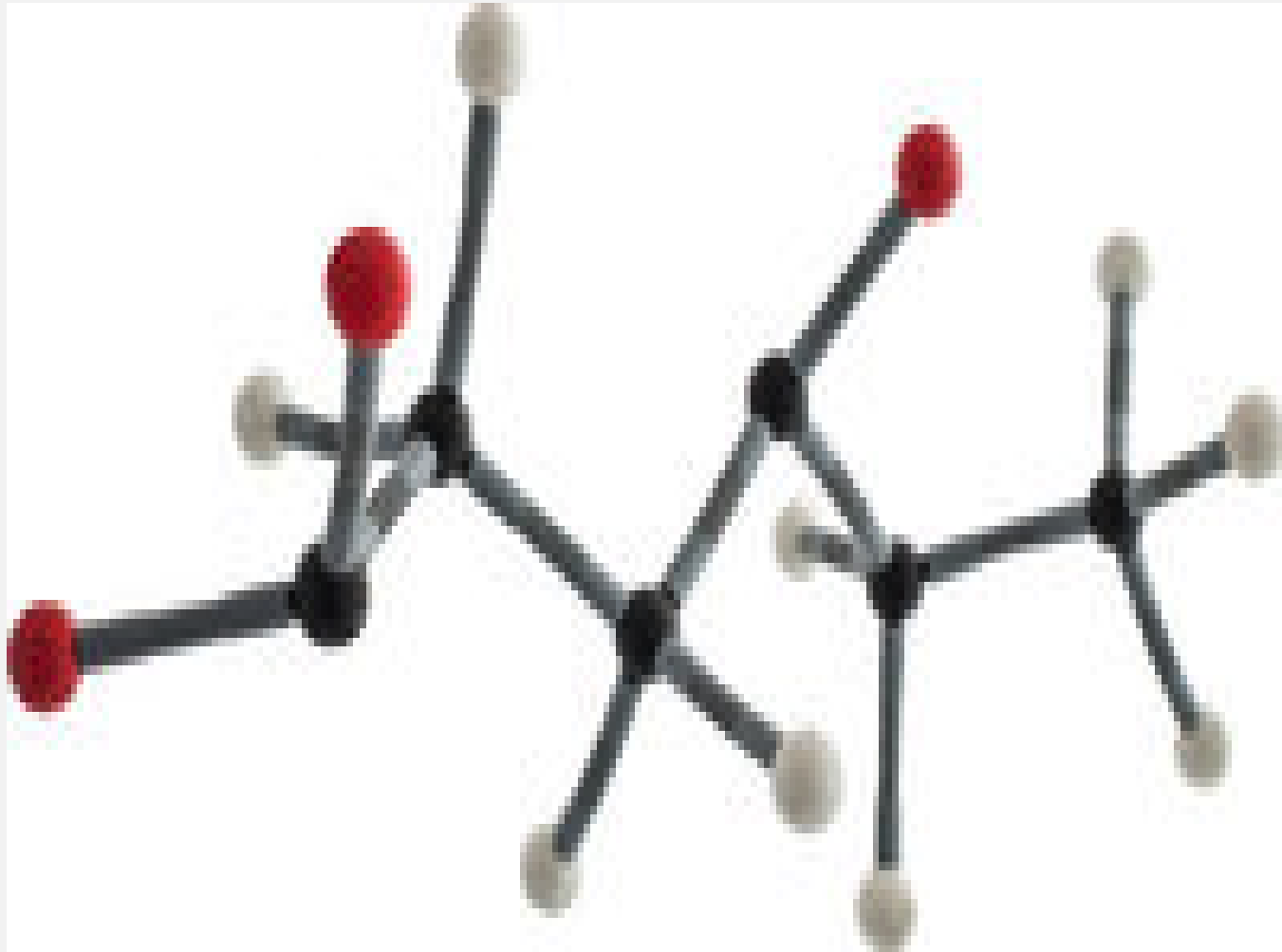
A Model Car



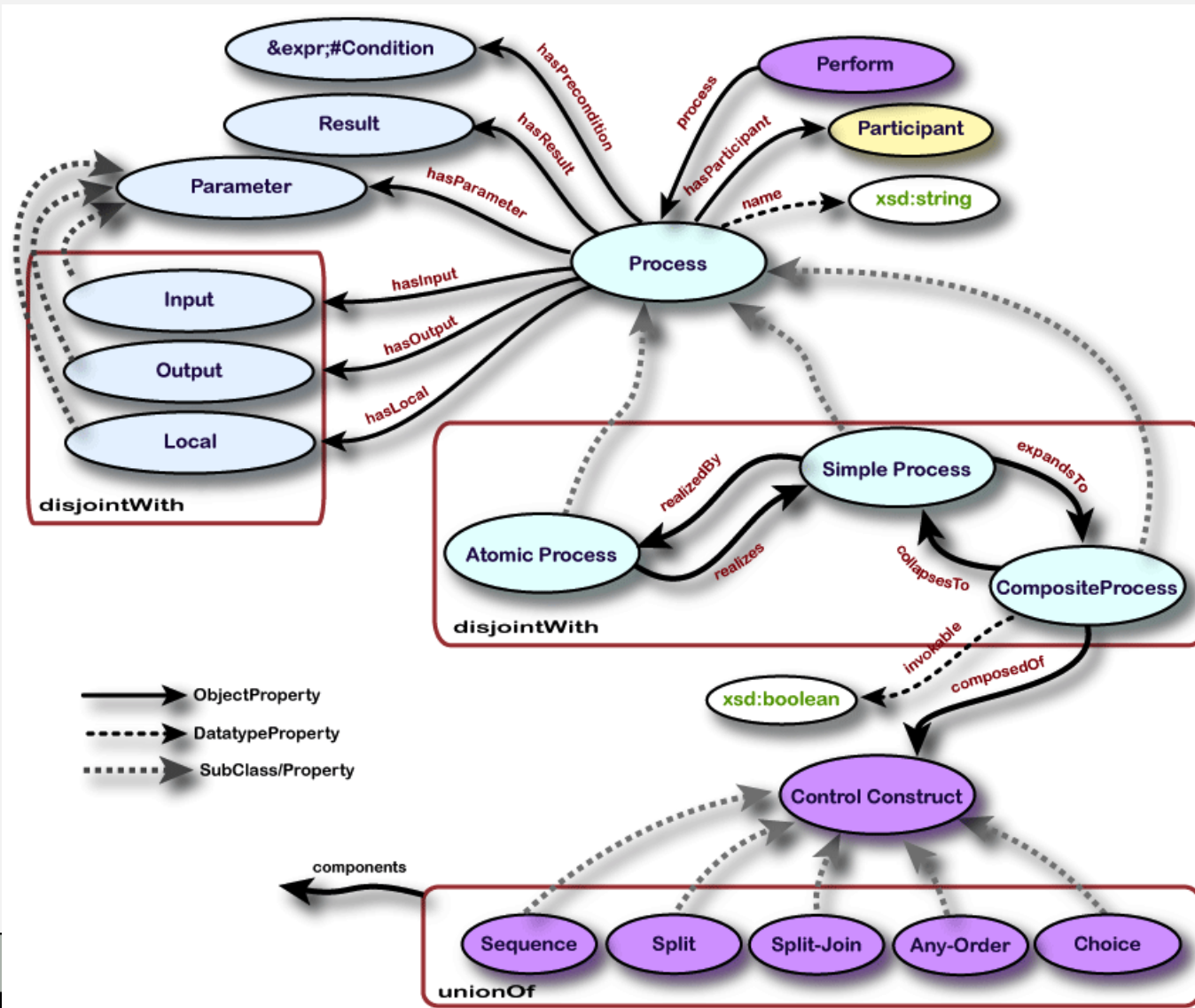
A Model Airplane (an Airplane Model?)

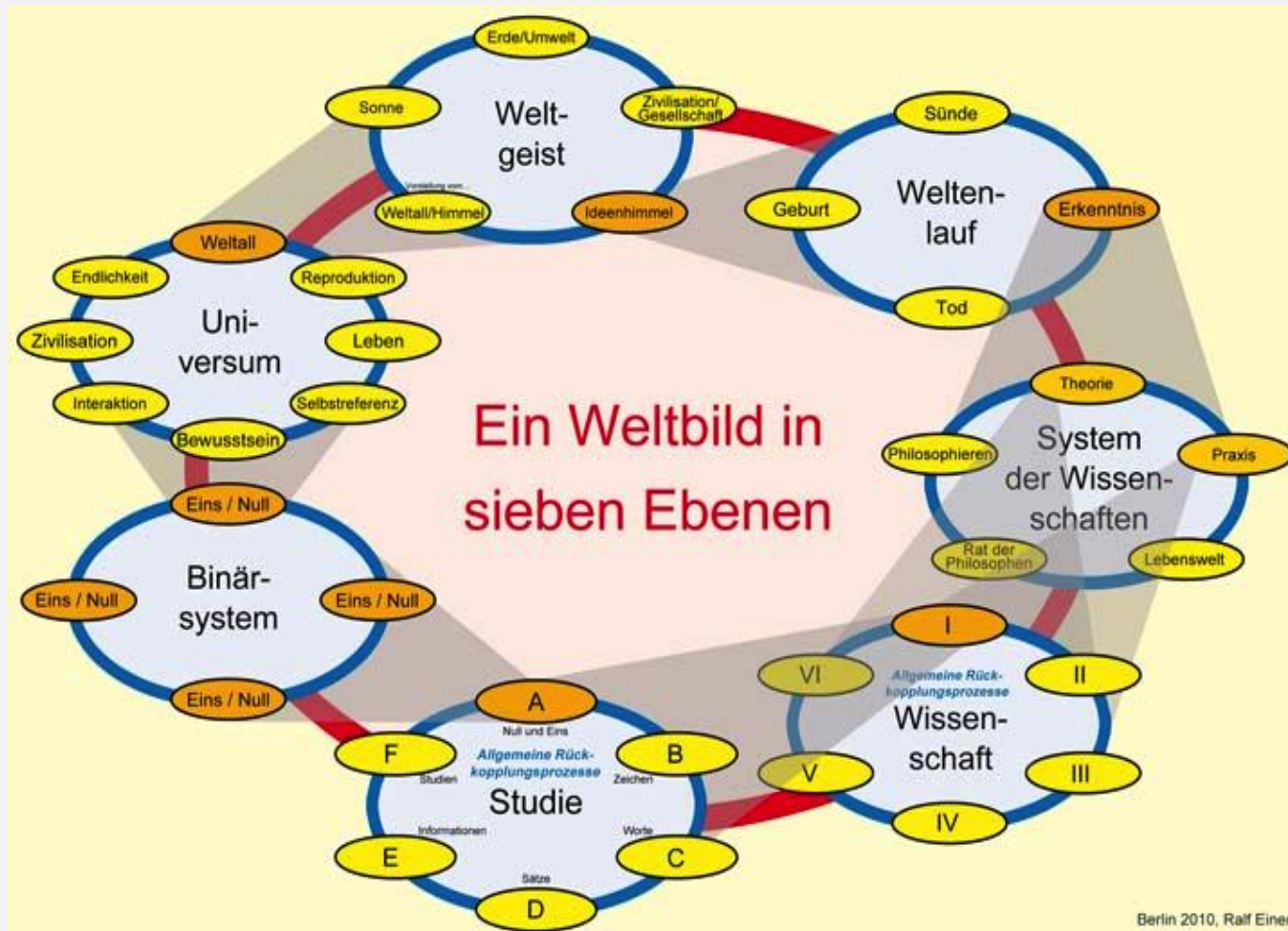


A 'Structural' Model



A Process Model

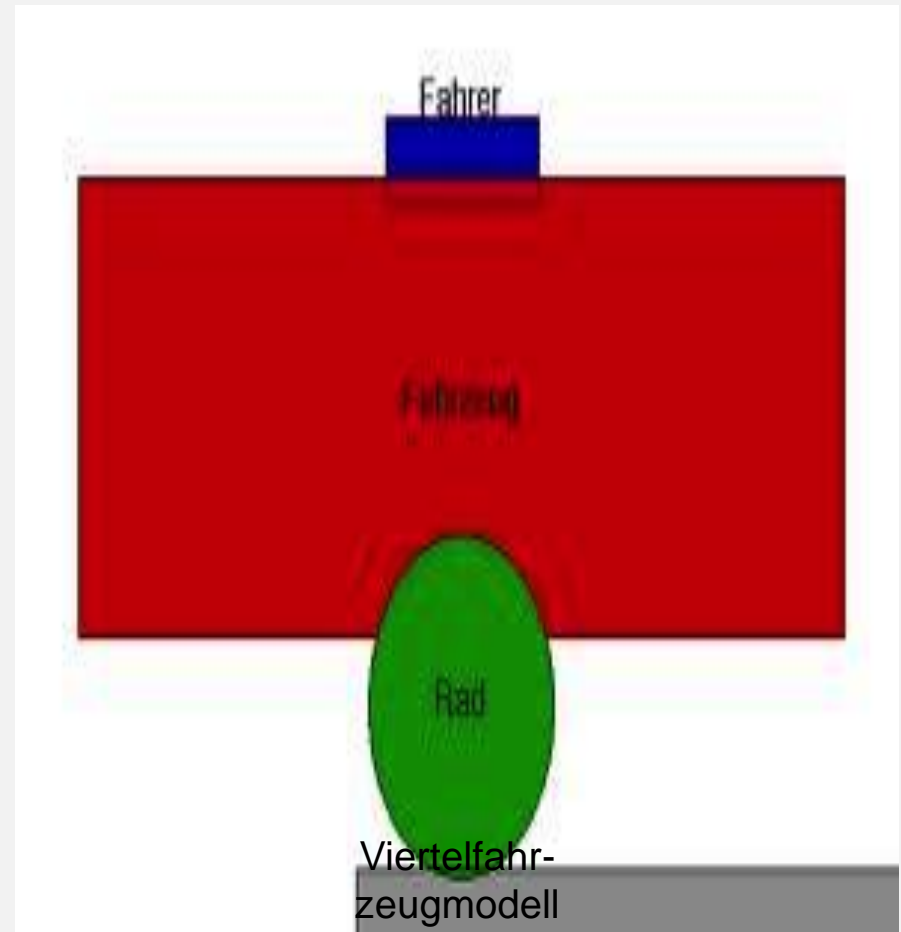




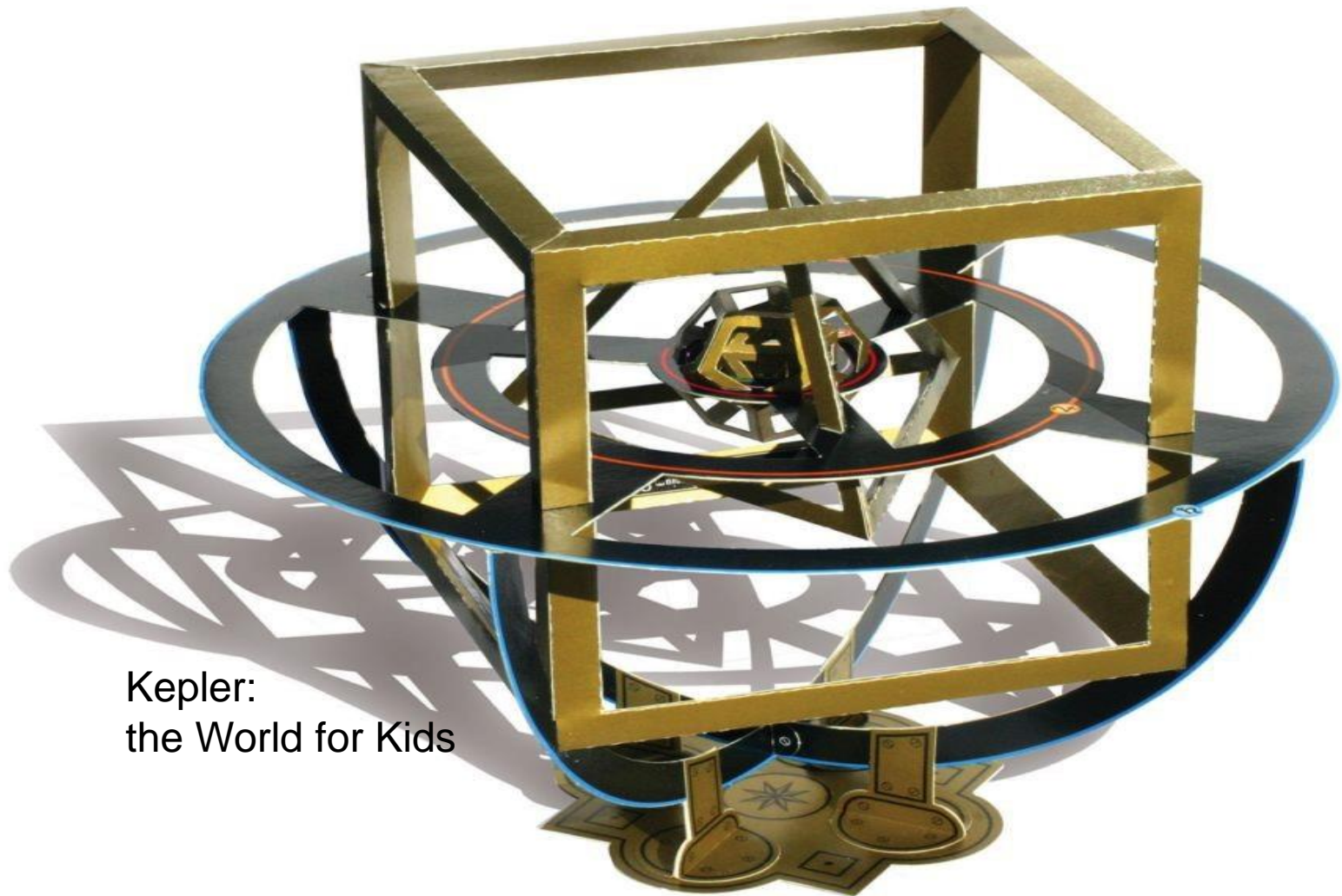
Model and Representation / Picturing



Model and Reduction / Abstraction

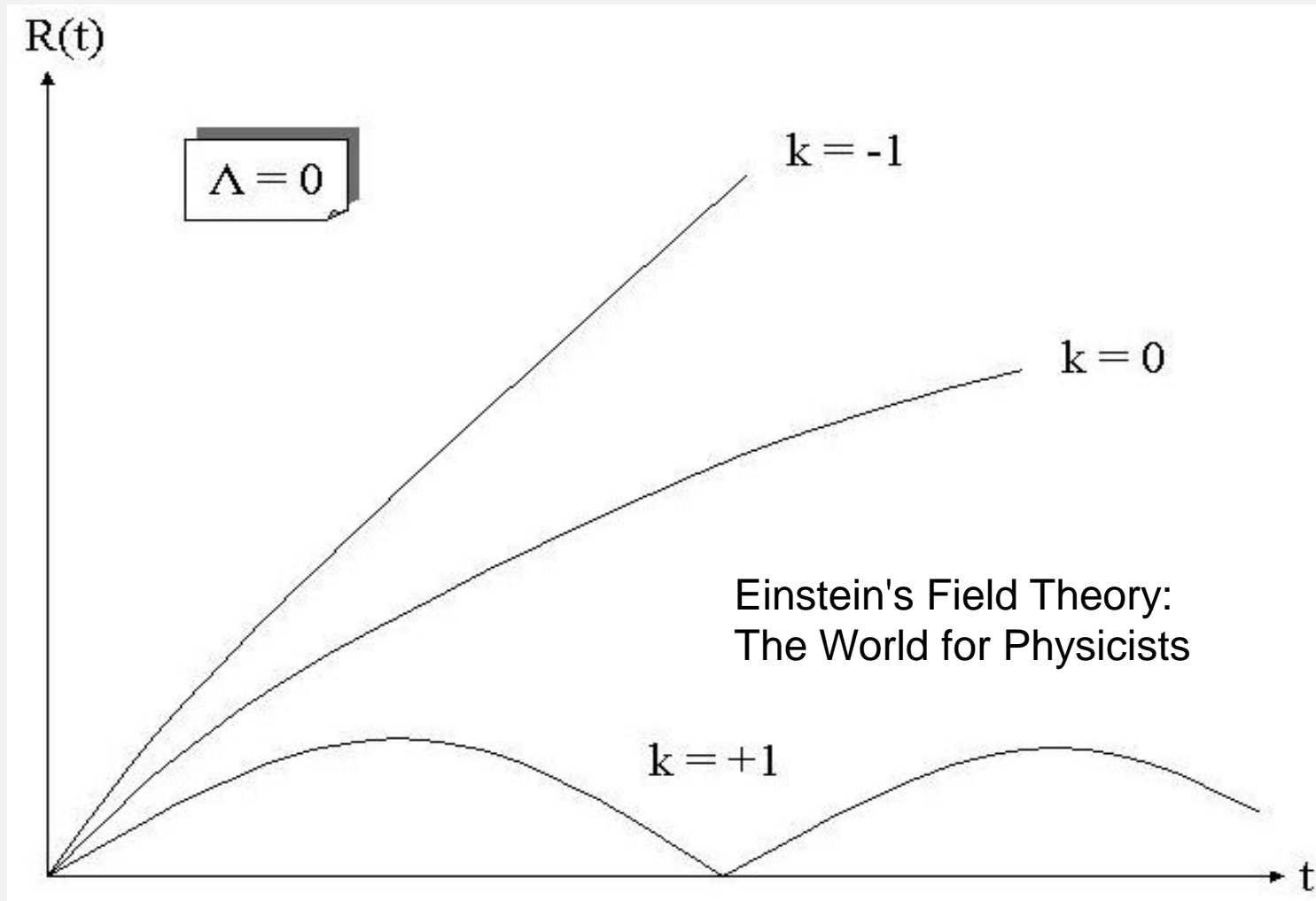


Pragmatic Orientation of Models (1)



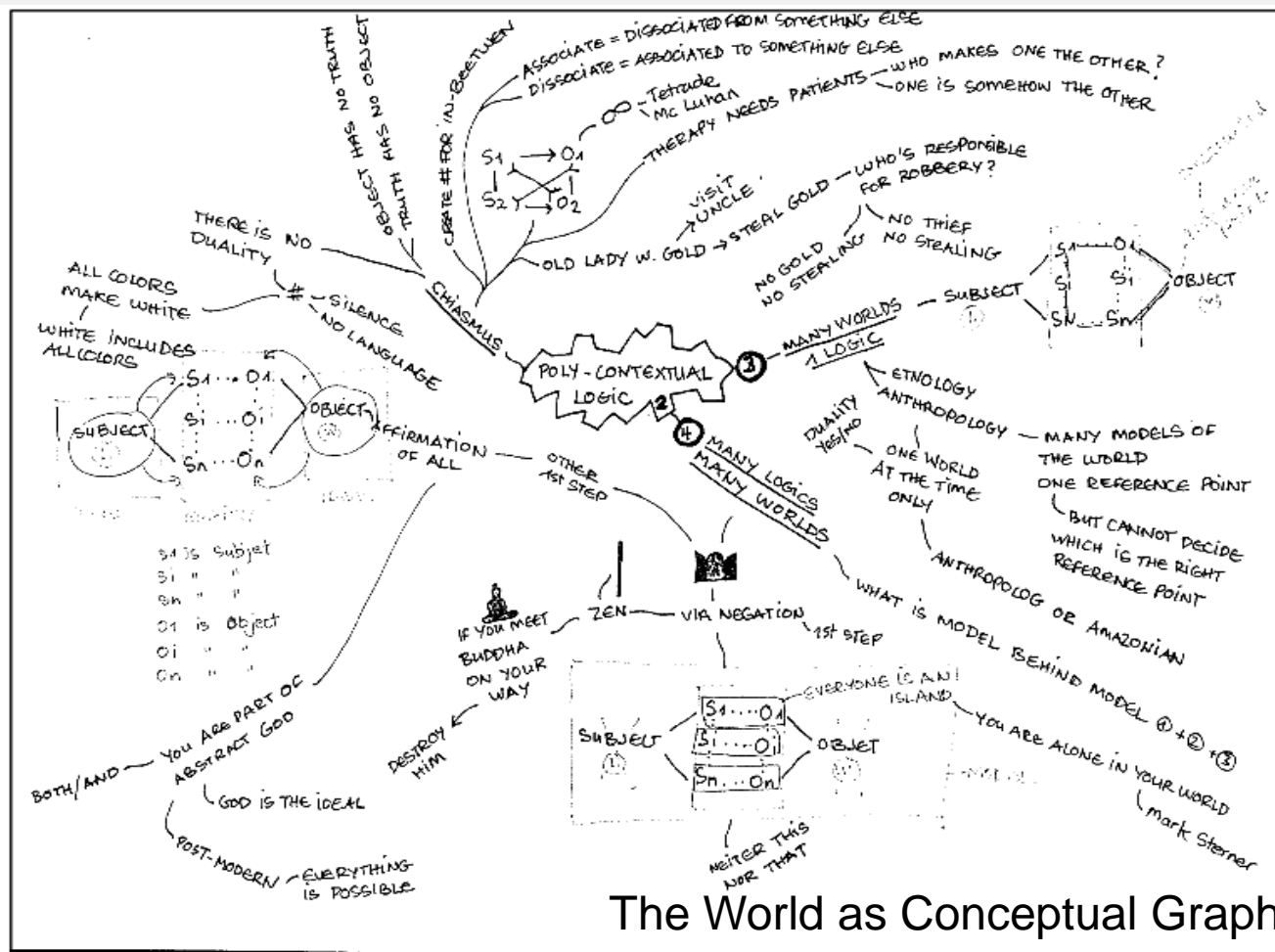
Kepler:
the World for Kids

Pragmatic Orientation of Models (2)



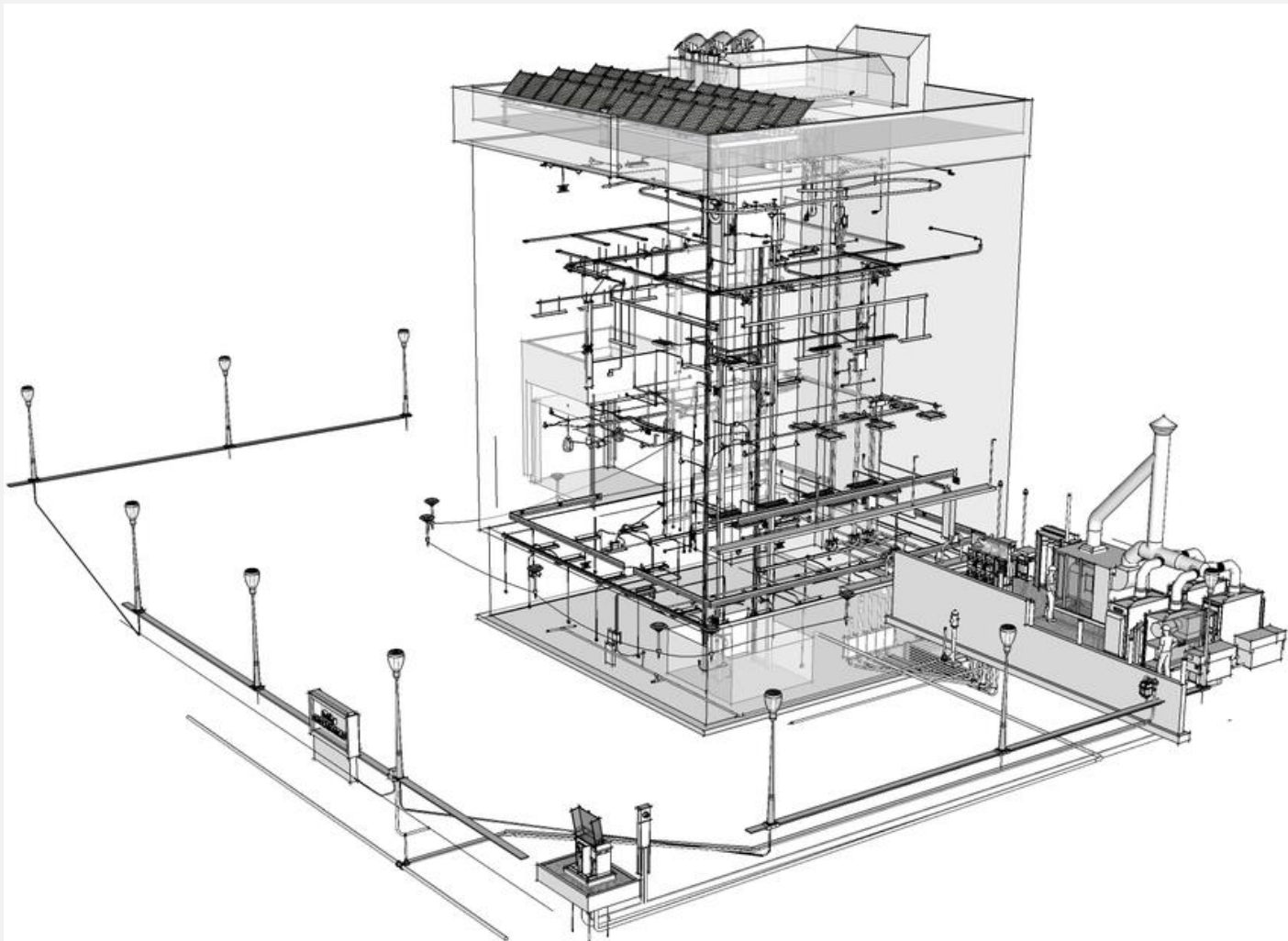
Einstein's Field Theory:
The World for Physicists

Pragmatic Orientation of Models (3)

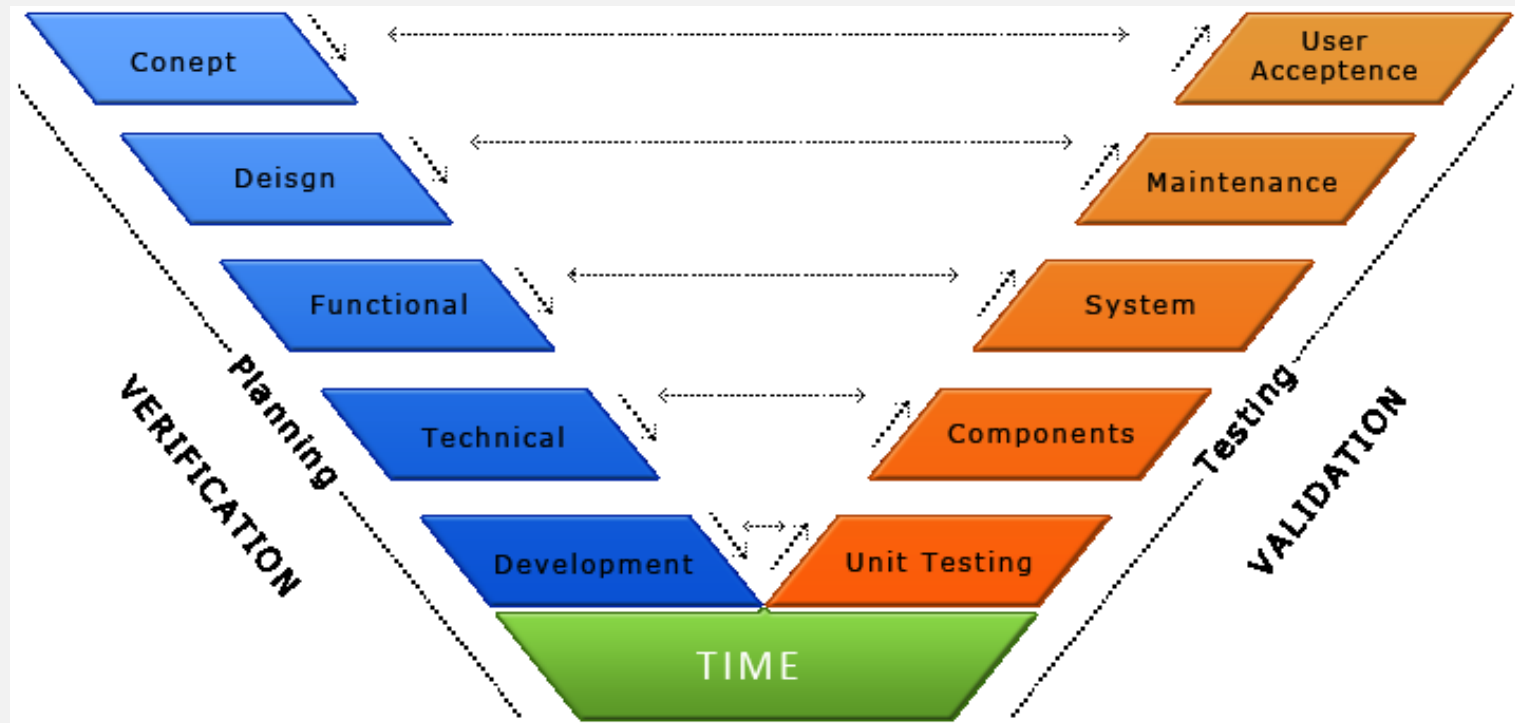


The World as Conceptual Graph

A Construction Model



The V-Model (1)



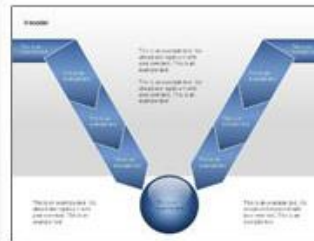
Many V-Models (2)



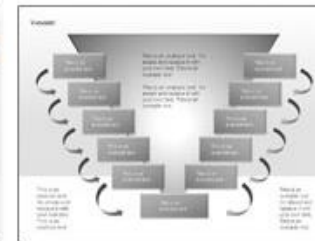
1



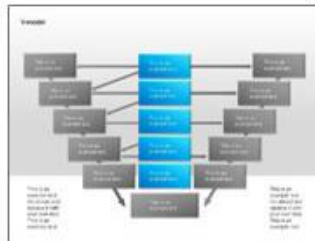
2



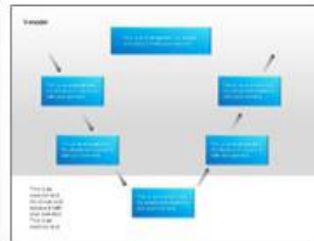
3



4



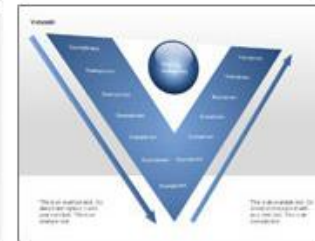
5



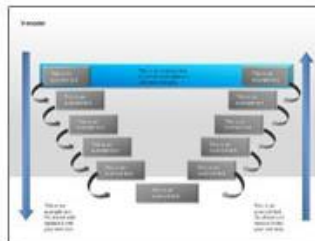
6



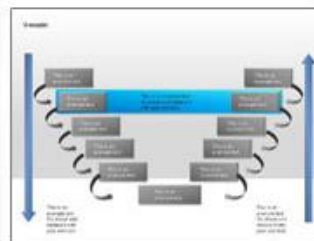
7



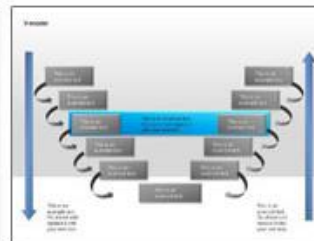
8



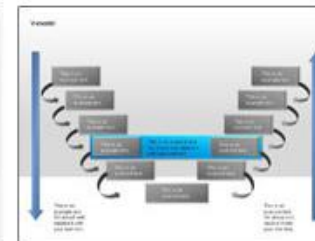
9



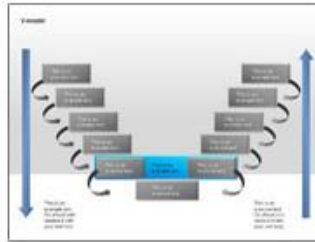
10



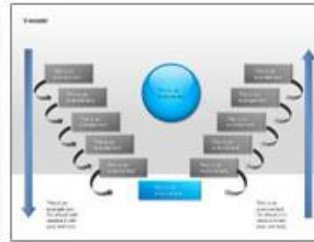
11



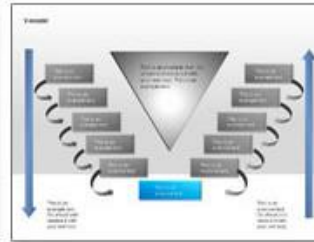
12



13

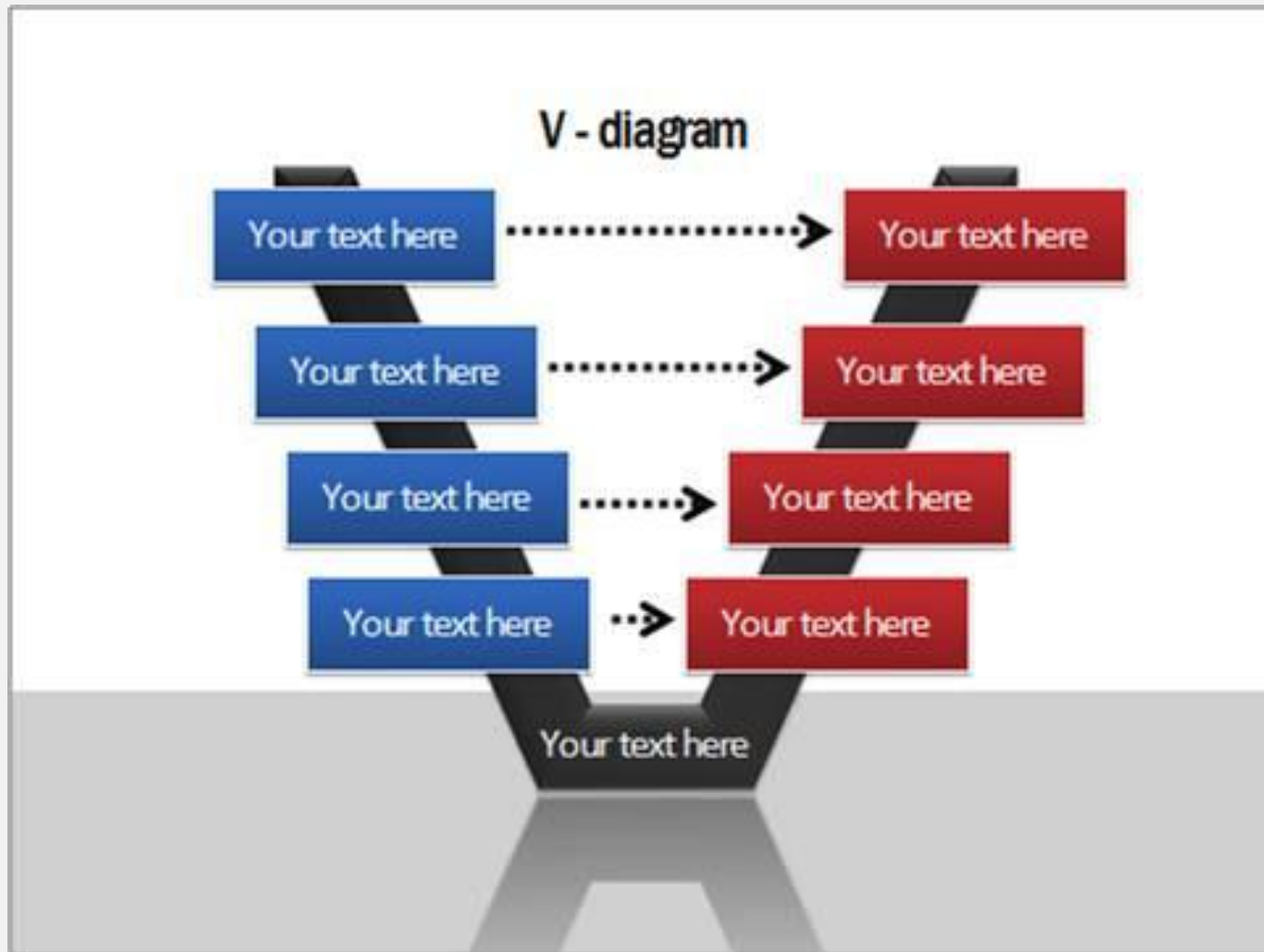


14



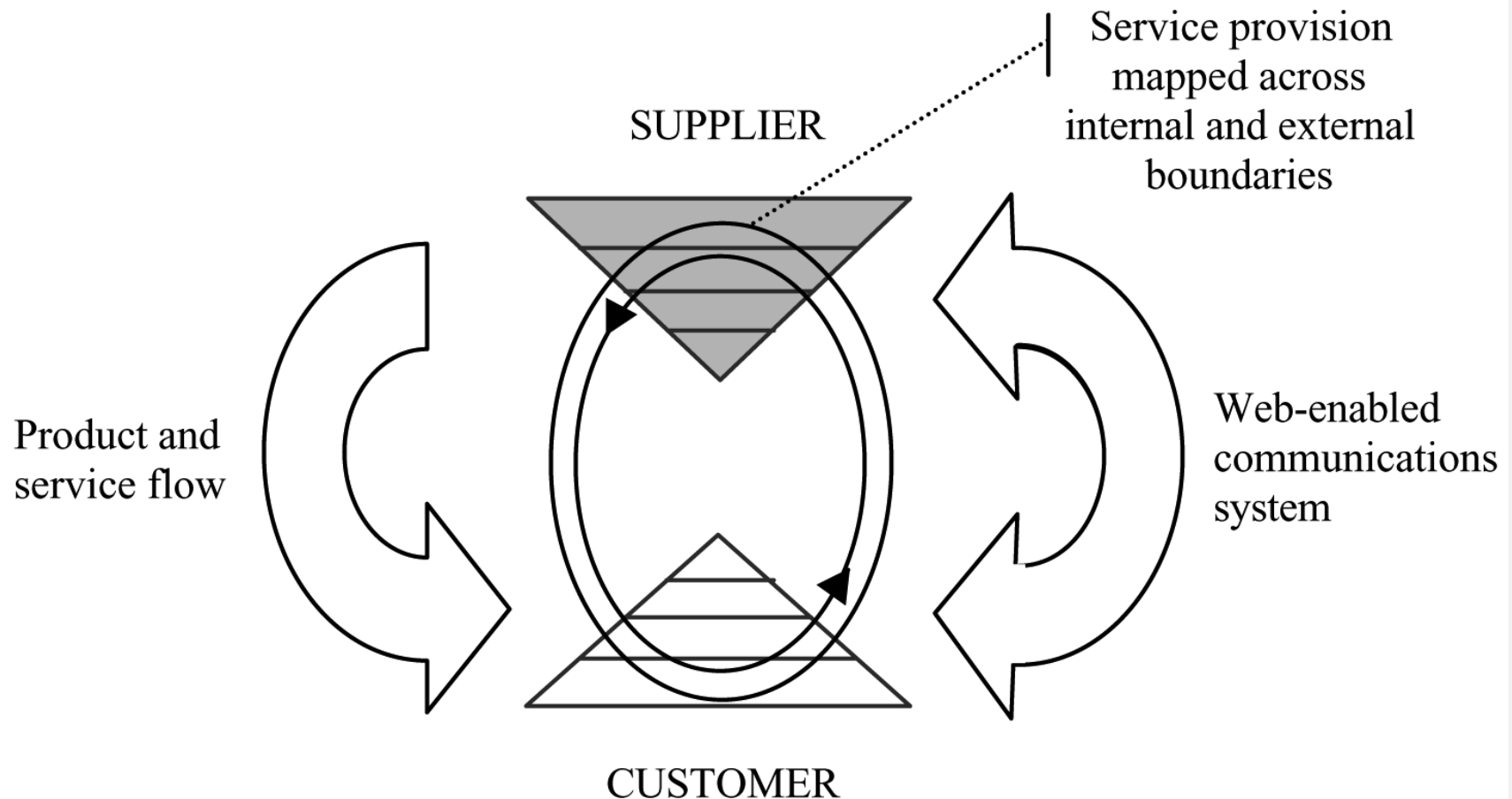
15

The Generic V-Model (3)



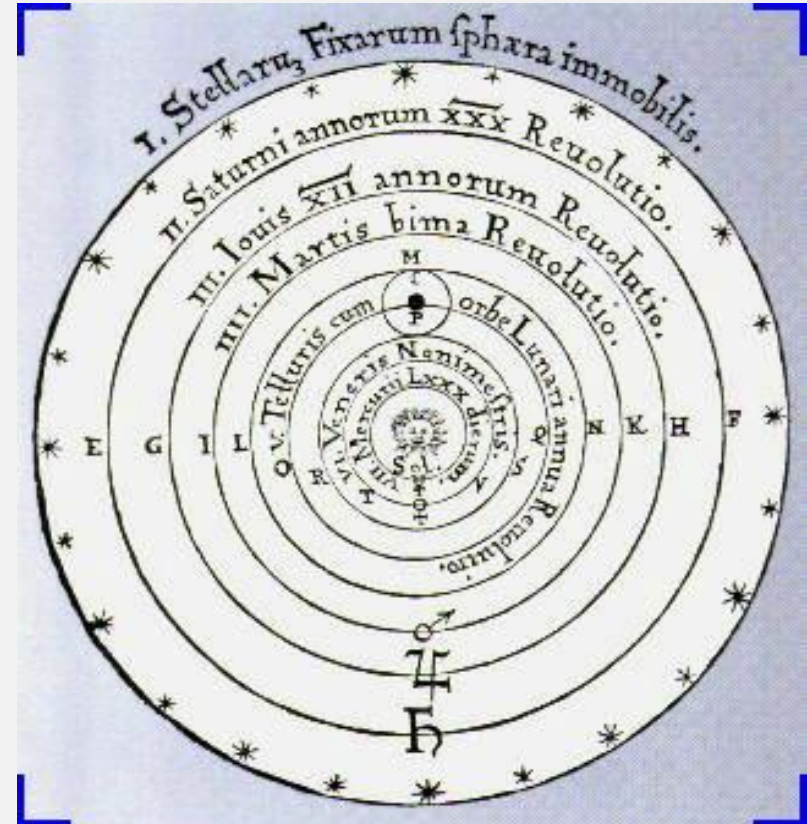
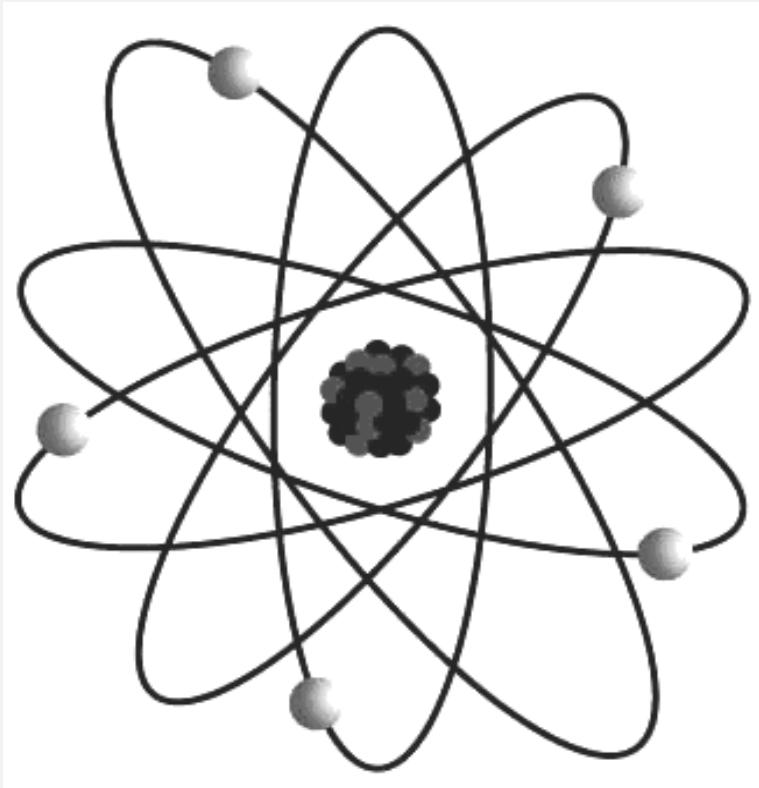
An Interaction Model

A Circle, two Triangles



Model vs. Metaphor: Atoms as modeled by Bohr – and the Universe as modeled by Copernicus

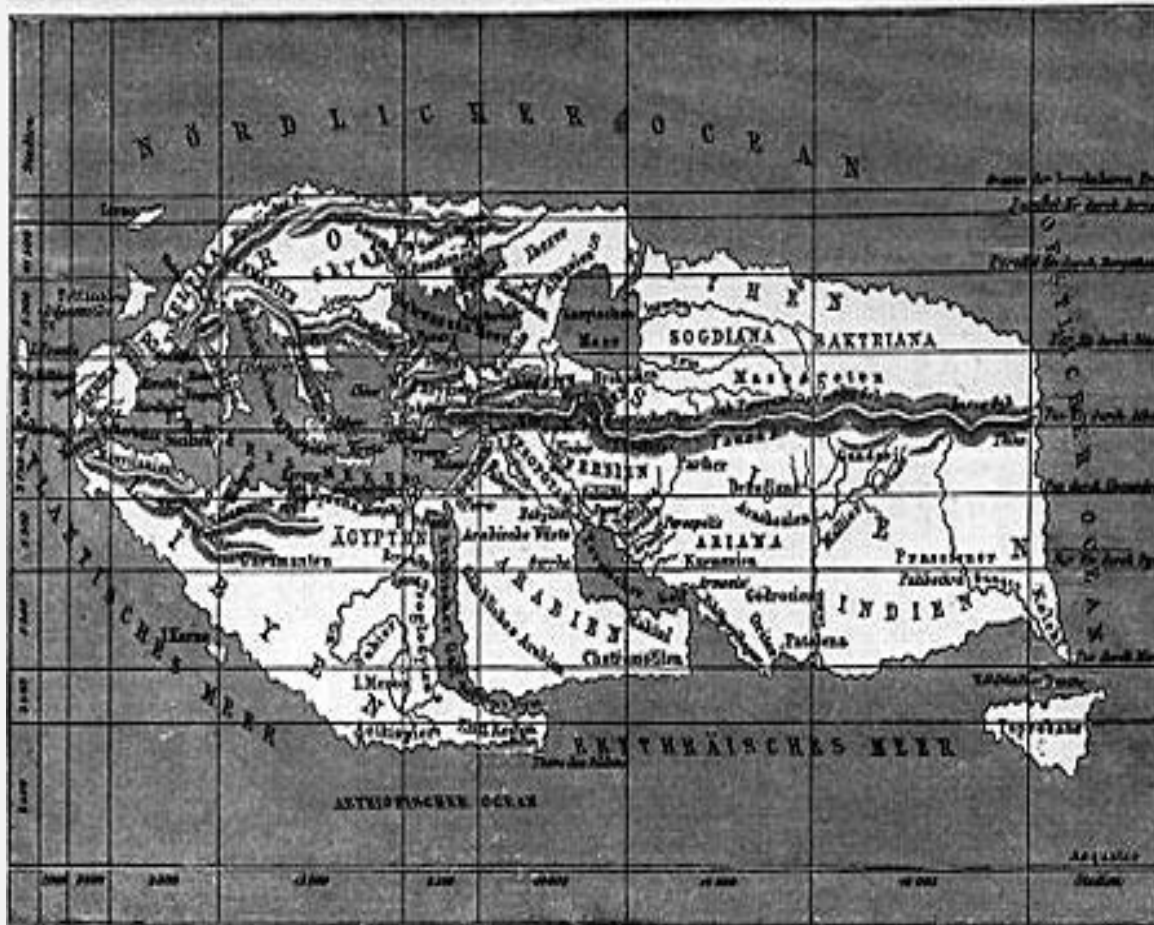
KU LEUVEN



A mythical World Model



Between Myth and Representation



From Myth to Metaphor



6 Statements on Models. 1 Question

- Models do not represent “reality”, but rather a perspective of 'world'.
- Models are instructions for creating 'reality'.
- Models are selective: they operate on reduction and abstraction.
- Reduction and abstraction always are intentionally guided.
- 'Powerful' models often operate with substantial metaphoric, connotative and/or symbolic 'surplus'.
- In this perspective, works of art are eminently powerful models!
- Which are our intentions when modeling the scholarly domain?

The Wittenstein Incubator

- a) Identify the intended functional extension of the 1st Pundit & Korbo versions (→ visualisation!)
- b) Stabilise scholarly domain model
- c) Identify additional specialisations of primitives
- d) Formalise, ontologically model such specialisations
- e) Populate the platform with Wittgenstein's Brown Book and related material
- f) Have ~10 scholars work in that environment

Wittgenstein SOURCE Privacy Credits

EXPLORE DOCUMENTATION ARCHIVE

About WittgensteinSource

SEARCH

ARCHIVE

Legend D N

Ms-114: X. Philosophische Grammatik m F D N

Ms-115: XI, Philosophische Bemerkungen m F D N

Ms-139a: Lecture on Ethics m F D N

Ms-140: Grosses Format m F D N

Ms-141 m F D N

Ms-148: C4 m F D N

Ms-149: C5 m F D N

Ms-150: C6 m F D N

Ms-152: C8 m F D N

Ms-153a: Anmerkungen m F D N

Ms-153b m F D N

Ms-154 m F D N

Ms-155 m F D N

Ms-156a m F D N

Ts-201a1: Notes on Logic m F D N

Ts-201a2: Notes on Logic m F D N

Ts-207: Lecture on Ethics m F D N

Ts-212 m F D N

Ts-213: Big Typescript m F D N

Ts-310: Brown Book m F D N

Ts-310 Facsimile

Ts-310,1 Ts-310,2

Ts-310,3 Ts-310,3a

Ts-310,4 Ts-310,5

Ts-310,6 Ts-310,7

WAB DISCOVERY (2008): Wittgenstein TS 310 G Trinity College Cambridge, Oxford University Press, University of Bergen

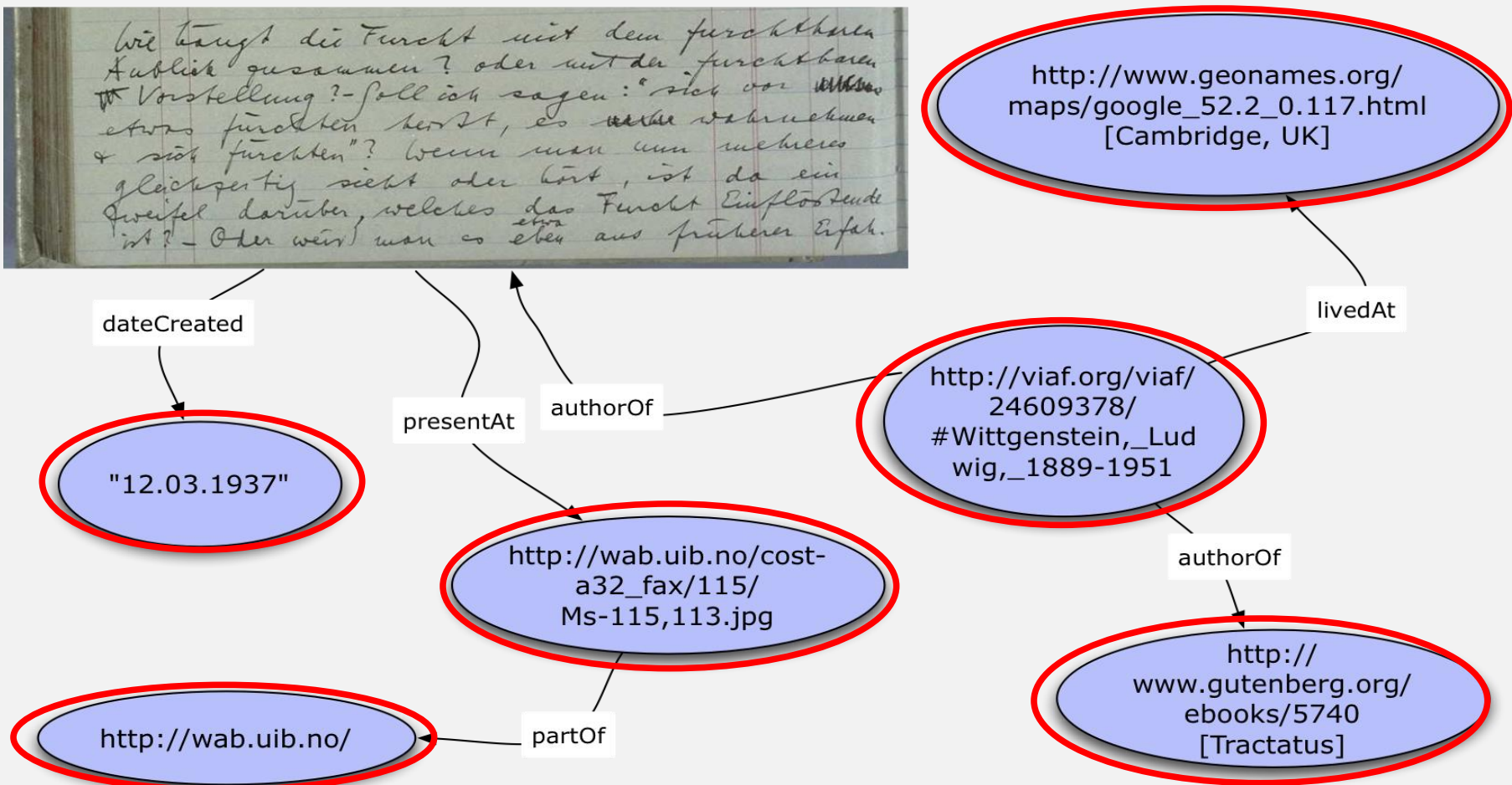
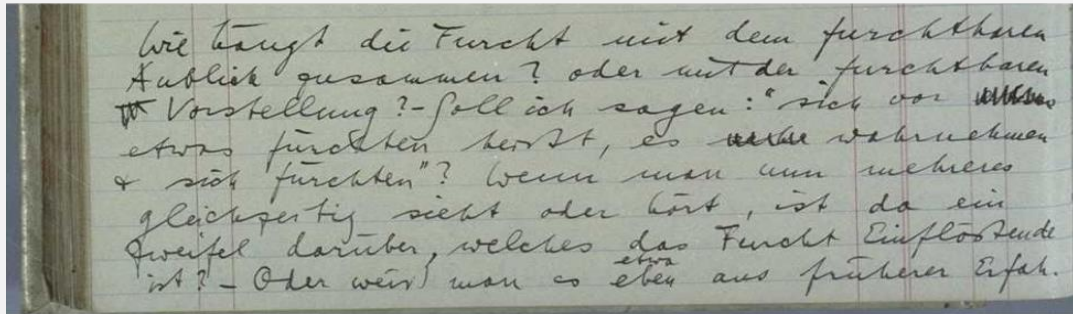
Reproduced by permission of the Master and Fellows of Trinity College, Cambridge (Trinity College Library), Oxford University Press, Oxford ("Wittgenstein's Nachlass: The Bergen Electronic Edition", 2000), the University of Bergen, Bergen, and Uni Research (earlier "Unifob"), Bergen (Wittgenstein Archives at the University of Bergen). The sale, further reproduction or use of this or any image on the site for commercial purposes without prior permission from the right holders is prohibited.

The screenshot displays the Pundit web application interface. The top navigation bar includes a 'Pundit' logo, a 'Help' button, and user information for 'Steffen Hennicke'. The main content area is divided into several sections:

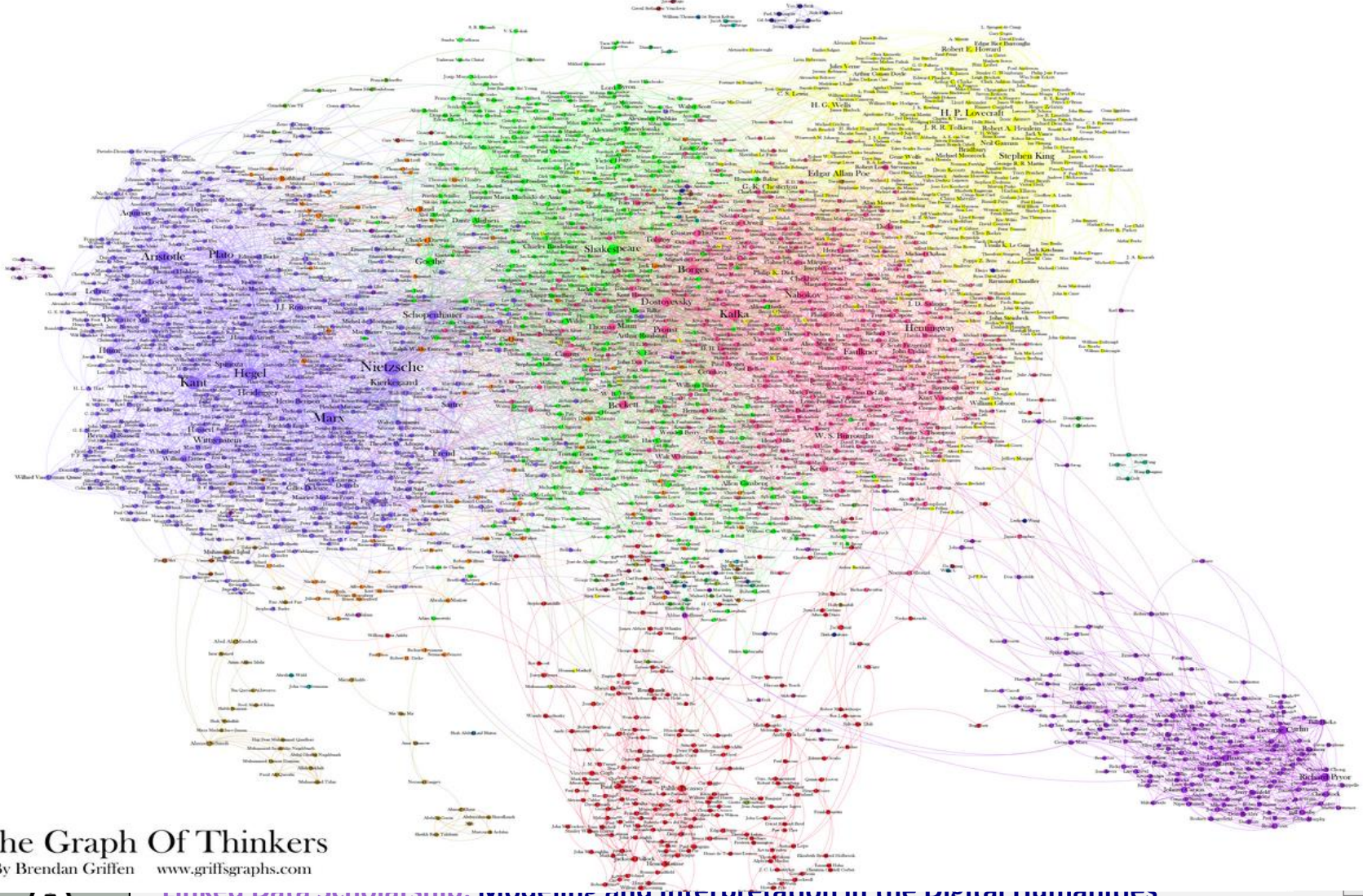
- PREVIEW:** Shows the text fragment 'Before I begin to speak about my subject proper let me say a few introductory words. I feel I will have great difficulties in communicating the thoughts which I want to communicate, to you & I want to mention some of these difficulties because I think that this may possibly diminish them. The first I will mention — but by no means the greatest — is that, as you... know, English is not my native language & my expression will therefore not be as clear & precise as it would be desirable when one has something very difficult to communicate. Please help me in my task of making myself understood by overlooking as much as possible the faults against English grammar which will constantly occur in my speech. The second difficulty which I will mention seems to me to be by far more serious & to explain it I must tell you why I have chosen the subject I have chosen. When your former secretary honoured me by asking me to read a paper to your society the first thought that came into my head was that I would certainly do it & the second was this: I said to myself that if I had the opportunity of talking to a room full of people that I would use this opportunity to say something that comes from my heart & not to misuse the time that I was given by explaining some scientific matter to you which to be properly explained would need a course of lectures or an audience specially trained in one particular line of thought & that I would still less misuse this opportunity of speaking to you by giving you a popular lecture, say on logic, which would serve to make you believe that you understand a thing which as a matter of fact you don't understand (& which it is not a bit necessary that you should) & to gratify the very lowest of modern desires viz. the superficial curiosity about the latest discoveries of scientists I decided — I say — that I should use this opportunity to speak to you not as a logician, still less as a cross between a scientist & a journalist but as a human being who tries to tell other human beings something which some of them might possibly find useful, I say useful not interesting. The third and last difficulty I will mention is one that adheres to most philosophical explanations & it is this that it sometimes is almost impossible to explain a matter in such a way that the hearer at once sees the road he is lead & the endgoal to which it leads. That is to say it so very often happens that the hearer thinks "I understand perfectly what he says but what on earth is he driving at" or else that he sees what one is driving at & thinks "that's all very well but how is he going to get there". This perhaps is the gravest difficulty & all I can do is to ask you to be patient & to hope that in the end we will see both the road & where it leads to. — Now let me begin.

The interface also features a 'Text-fragment' tab, a 'Triple Composer' for creating semantic triples, and a 'Metadata' section on the right side. The metadata section includes fields for 'Created by', 'On', 'ID', and 'Types'.

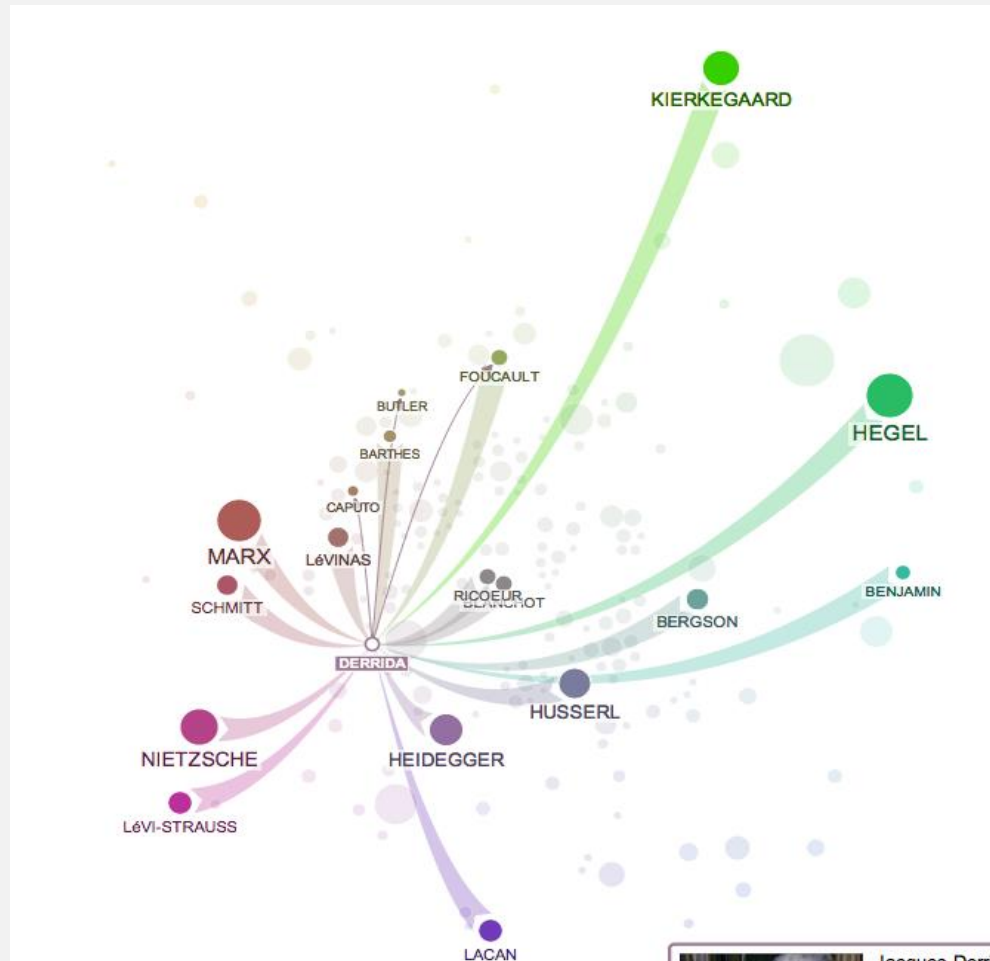
Contextualising Wittgenstein



Visualising: Graph of Thinkers



Visualising: Graph of Philosophers



Nonlinear Reading

<http://texttexture.com/index.php>

- A **Social Semantic Scholarly Graph**
 - Enabling **interaction** (via Pundit and Edgemaps/LODLive)
 - Enabling **heuristic operations** (building on RDFS inference)
 - As an **object of scholarly study** (graph evolution modeling requires named graph based extensions for **versioning**, **provenance**, **authorisation** et.)
- **Ontology components** for modeling scholarly discourse and interaction
 - Beware: “The limits of my language mean the limits of my world.” Tractatus, 5.6
- And, most importantly: “Whereof one cannot speak, thereof one must be silent.” Tractatus, 7

Bibliography (1)

- Anderson, Sheila; Blanke, Tobias; Dunn, Stuart (2010): Methodological commons: arts and humanities e-Science fundamentals. In: Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 368 (1925), S. 3779–3796.
- Atkins, Daniel. E., et al. (2003) Revolutionizing Science and Engineering Through Cyberinfrastructure. Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.
= <http://www.nsf.gov/od/oci/reports/atkins.pdf>
- Bamboo (2010): Project Bamboo Scholarly Practice Report.
= <http://www.projectbamboo.org/wp-content/uploads/Project-Bamboo-Scholarly-Practices-Report.pdf>
- Benardou, Agiatis; Constantopoulos, Panos; Dallas, Costis; Gavrilis, Dimitris (2010): A Conceptual Model for Scholarly Research Activity. IConference 2010. = <https://www.ideals.illinois.edu/handle/2142/14945>
- Blanke, T., & Hedges, M. (2011). Scholarly primitives: Building institutional infrastructure for humanities e-Science. Future Generation Computer Systems. doi:10.1016/j.future.2011.06.006

Bibliography (2)

- Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure and the Internet*. Cambridge, Mass.: MIT Press.
- Brockman, William S.; Neumann, Laura; Palmer, Carole L.; Tidline, Tonyia J. (2001): *Scholarly Work in the Humanities and the Evolving Information Environment*. Washington, D.C.: Council on Library and Information Resources.
- Bush, Vannever. 'As We May Think'. *Atlantic Magazine* (July 1945). = <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>.
- Dörk, Marian; Carpendale, Sheelagh; Williamson, Carey (2011): Visualizing explicit and implicit relations of complex information spaces. *Information Visualization* 2012 11: 5. DOI: 10.1177/1473871611425872
- Doerr, M., Kritsotaki, A., & Boutsika, K. (2011). Factual argumentation—a core model for assertions making. *Journal on Computing and Cultural Heritage*, 3(3), 1–34. doi:10.1145/1921614.1921615
- Gradmann, S. (2010). *Knowledge = Information in Context : on the Importance of Semantic Contextualisation in Europeana*.
= <http://www.scribd.com/doc/32110457/Europeana-White-Paper-1>

Bibliography (3)

- Gradmann, S., & Meister, J. C. (2008). Digital document and interpretation: re-thinking “text” and scholarship in electronic settings. *Poiesis Praxis*, 5(2), 139–153. doi:10.1007/s10202-007-0042-y
- Johannessen, Harald (2011): *Debatt og argumentasjon: En innføring*. Oslo: Spartacus and Scandinavian Academic Press
- McCarty, Willard; Short, Harold (2002): *Mapping the Field*. Report of ALLC meeting held in Pisa, April 2002. = <http://www.allc.org/node/188>
- Palmer, C. L. (2000). *Configuring Digital Research Collections around Scholarly Work*. Paper presented at Digital Library Federation Forum, November 19, Chicago, Illinois.
- Palmer, C., Tefteau, L., & Pirmann, C. (2009). *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. Report. Development.
= <http://www.oclc.org/resources/research/publications/library/2009/2009-02.pdf>
- Rockwell, G. (2010, May 14). *As Transparent as Infrastructure: On the research of cyberinfrastructure in the humanities*.
= <http://cnx.org/content/m34315/1.2/>

- schraefel, m. c. (2007). What is an Analogue for the Semantic Web and Why is Having One Important? Manchester: ACM Hypertext 2007.
= <http://eprints.soton.ac.uk/264274/1/schraefelSWAnalogueHT07pre.pdf>
- Unsworth, J. (2000). Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? Symposium on Humanities Computing formal methods experimental practice. =
<http://www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html>
- Unsworth, J., et al. (2006). Our Cultural Commonwealth. Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences.
= <http://www.acsls.org/cyberinfrastructure/cyber.htm>
- Unsworth, John; Tupman, Charlotte (2011): Interview with John Unsworth, April 2011, carried out and transcribed by Charlotte Tupman. In: Marilyn Deegan und Willard McCarty (Hg.): Collaborative research in the digital humanities. Farnham: Ashgate, S. 231–239.

Thank you for your patience and attention

The Slides I did not Prepare

- Digital? Humanities??
- As opposed to what?
- Which purpose do we need this term for?
- ... (→ Discussion)

Thank you for your patience and attention